PROCEEDINGS

ICIST 2015

5TH INTERNATIONAL CONFERENCE ON INFORMATION SOCIETY AND TECHNOLOGY

KOPAONIK 8. - 11. MARCH 2015.

ICIST 2015

5th International Conference on Information Society and Technology

Proceedings

Publisher: Society for Information Systems and Computer Networks

Editors: Zdravković, M., Trajanović, M., Konjović, Z.

ISBN:978-86-85525-16-2

Issued in Belgrade, Serbia, 2015.

ICIST 2015
5th International Conference on Information Society and Technology (ICIST 2015)
International Programme Committee

Carlos Agostinho, UNINOVA, Portugal
Phil Archer, W3C/ERCIM, France
Alexis Aubry, Université de Lorraine, France
Miloš Bogdanović, University of Niš, Serbia
Osiris Canciglieri, Pontifícia Universidade Católica do Paraná, Brasil
David Chen, University Bordeaux 1, France
Carlos Coutinho, Universidade Nova de Lisboa, Portugal
Žarko Čojbašić, University of Niš
Mariana Damova, Mozaika, Bulgaria
Igor Dejanović, University of Novi Sad
Neven Duić, University of Zagreb, Croatia
Anna Fensel, Semantic Technology Institute (STI) Innsbruck, University of Innsbruck, Austria
Nataša Gospić, University of Belgrade, Serbia
Stevan Gostojić, University of Novi Sad
Elissaveta Gourova, Sofia University "St. Kliment Ohridski", Bulgaria
Wided Guédria, CRP Henri Tudor, Luxembourg
Irena Holubová, Faculty of Mathematics and Physics, Charles University, Czech Republic
Daniel Hladky, Ontos AG/National Research University – Higher School of Economics, Switzerland/Russia
Dragan Ivanović, University of Novi Sad
Valentina Janev, Institute Mihajlo Pupin, Serbia
Cecil Joe, Oklahoma State University, USA
Dorina Kabakchieva, University of National and World Economy, Sofia, Bulgaria
Martin Kaltenböck, Semantic Web Company GmbH, Austria
Zora Konjović, University of Novi Sad, Serbia
Nikola Korunović, University of Niš, Serbia
Aleksandar Kovačević, University of Novi Sad
Srđan Krčo, University of Belgrade, Serbia
Lea Kutvonen, University of Helsinki, Finland
Jens Lehmann, Institute of Computer Science, University of Leipzig, Germany
Vuk Malbaša, University of Novi Sad
Zoran Marjanović, University of Belgrade, Serbia
Miloš Madić, University of Niš, Serbia
Istvan Mezgar, Computer and Automation Research Institute, Hungarian Academy of Sciences, Hungary
Dragan Mišić, University of Niš, Serbia
Branko Milosavljević, University of Novi Sad
Gordana Milosavljević, University of Novi Sad
Néjib Moalla, University Lyon 2 Lumière, France

Arturo Molina Gutiérrez, Tecnológico de Monterrey, Mexico
Ovidiu Noran, Griffith University, Australia
Đorđe Obradović, University of Novi Sad
Dušan Okanović, University of Novi Sad
Hervé Panetto, Université de Lorraine, France
Milan Paroški, University of Novi Sad
Kostas Patroumpas, School of Electrical and Computer Engineering - N.T.U.A., Greece
Valentin Penca, University of Novi Sad
Michaël Petit, Université de Namur, Belgium
Erik Proper, CRP Henri Tudor, Luxembourg
Eduardo Rocha Loures, Pontifícia Universidade Católica do Paraná, Brasil
David Romero, Tecnológico de Monterrey, Mexico
Camille Salinesi, Pantheon-Sorbonne University, Computer Science Research Center (CRI), France
Joao Sarraipa, UNINOVA, Portugal
Goran Savić, University of Novi Sad
Milan Segedinac, University of Novi Sad
Jean Simão, Universidade Tecnológica Federal do Paraná, Brasil
Goran Sladić, University of Novi Sad
Jelena Slivka, University of Novi Sad
Marten van Sinderen, University of Twente, Netherlands
Richard Mark Soley, OMG, USA
Kamelia Stefanova, Faculty of Applied Informatics and Statistics, University of National and World Economy, Sofia, Bulgaria
Leonid Stoimenov, University of Niš, Serbia
Anderson Szejka, Pontifical University Catholic of Paraná, Brasil
William Wei Song, Dalarna University, Sweden
Miroslav Trajanović, University of Niš, Serbia
Milan Trifunović
Bruno Vallespir, Université Bordeaux 1 / IMS, France
Milan Vidaković, University of Novi Sad
Nikola Vitković, University of Niš, Serbia
Sanja Vraneš, Institute Mihajlo Pupin, Serbia
Georg Weichhart, Johannes Kepler Universität Linz, Austria
Miroslav Zarić, University of Novi Sad
Jelena Zdravković, Stockholm University, Sweden
Milan Zdravković, University of Niš, Serbia
Martin Zelm, INTEROP-VLab, Belgium

CONTENT

Towards the Future Internet: A Foreword to the Proceedings of the 5th International Conference on Information Society and Technology

Milan Zdravković*, Miroslav Trajanović*, Zora Konjović**
* Laboratory for Intelligent Production Systems (LIPS),
Faculty of Mechanical Engineering, University of Niš, Niš, Serbia
** Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia

milan.zdravkovic@gmail.com, miroslav.trajanovic@masfak.ni.ac.rs, ftn_zora@uns.ac.rs

I. INTRODUCTION

The research on the Future Internet is driven by need to overcome some of the shortcomings of the current protocols and architectures, mostly in terms of performance, reliability, scalability, security, as well as the other categories. In addition, some societal, economic and business aspects are also considered in defining the long term vision for the Future Internet.

Based on the huge interest of the research community in this topic, which is also supported by the commitment of the European Commission to fund research in this area in scope of Horizon 2020 program, the ICIST Organizational Committee (OC) decided to make effort towards the further promotion and motivation in this exciting area. Hence, this year's edition of the conference was focused on the topic: Future Internet: Technologies, Architectures and Applications.

The 5th International Conference on Information Society and Technologies was organized in Kopaonik, Serbia, 8-10.3.2015. Despite the extreme travel conditions at that time, the conference gathered more than 300 participants from all over the world, to discuss one the recent research results.

The conference was supported by the International Program Committee (IPC), with researchers from 20 countries, namely Australia, Austria, Belgium, Brazil, Bulgaria, Czech Republic, Finland, France, Germany, Greece, Hungary, Luxembourg, Mexico, Netherlands, Portugal, Serbia, Sweden, Switzerland, United Kingdom and United States.

II. SCIENTIFIC PROGRAMME OF 5TH INTERNATIONAL CONFERENCE ON INFORMATION SOCIETY AND TECHNOLOGIES

ICIST 2015 received 114 submissions. The number of submissions is considered as an important evidence of the sustainable conference development and significant growth trend, after 90 submissions received last year. The international character of the conference is again demonstrated by the fact that authors or co-authors of the submitted papers were affiliated to the research institutions from 26 countries.

Based on the outcomes of the evaluation process (each paper is reviewed by 1-3 members of IPC), 64 papers were invited to be presented in some of the regular or special sessions. In addition, 34 papers were invited to be presented in two poster sessions.

The conference hosted three distinguished keynote speakers, namely: Dr. Richard Mark Soley, Chairman and CEO, OMG, USA; Vladimir Weinstein, Engineer Manager, Google, USA; Orri Erling, Program Manager for the Virtuoso Hybrid RDBMS, OpenLink Software.

A. Scientific sessions

During the preparation of the conference, a list of relevant topics was made by the OC and selected researchers were invited to organize the special sessions on these topics. Based on the response, it was decided that ICIST 2015 will host four special sessions, namely: Model-Based information Systems Engineering (MBiSE), Open Data and GIS applications (ODaGIS), ICT in Biomedical Engineering (ICTiBE) and Next-Generation Enterprise Information Systems (NGEIS).

Other papers, submitted and accepted for presentation in the regular program were classified in the following sessions: Software Development, Energy, Environment and Sustainable Development, Simulation and Optimization, Information Systems and Information Society. All accepted papers, presented in regular or special sessions are published in Volume 1 of this book.

Besides these topics, the papers presented at the poster sessions addressed also the area of learning management systems. All accepted papers, presented at the poster sessions are published in Volume 2 of this book.

1) Special Sessions

After significant response and large number of received submissions, ICTiBE session was organized in two timeslots, with 14 papers. Session was organized and chaired by Osiris Canciglieri Junior, Pontifical Catholic University of Paraná, Brasil, Nenad Filipović, University of Kragujevac, Serbia and Miroslav Trajanović, University of Niš, Serbia. The presented papers addressed different issues of modeling medical devices and anatomy; diagnosis, observations and assessment, based on biofeedback apps, neural networks and data mining algorithms; and developing systems for knowledge discovery and medical device design.

Organized by Hervé Panetto, Université de Lorraine, France, Richard Mark Soley and Milan Zdravković,

MBiSE presented seven papers. The authors dealt with runtime models, modeling requirements, document management systems, airport management systems, supply chain management systems and semantic conflict resolution.

Third special session on Open Data and GIS Applications was chaired and organized by Phil Archer, W3C Data Activity Lead, W3C/ERCIM, France, Jens Lehmann, University of Leipzig, Germany, and Valentina Janev, Institute "Mihajlo Pupin", Belgrade, Serbia. The session presented work related to the issues of open satellite data, spatio-temporal analysis and visualization of linked data, collaboration between public administrations, visual analytics of traffic-related open data and benchmarking of geospatial systems.

NGEIS session addressed the different issues of the Next-Generation Enterprise Information System, the concept that has been recently elaborated by the IFAC TC5.3 Technical Committee for Enterprise Integration and Networking of International Federation for Automatic Control, one of the supporters of the conference. Organized and chaired by Ricardo Jardim-Gonçalves, UNINOVA, Portugal, and Elisaveta Gourova, Sofia University "St. Kliment Ohridski", Bulgaria, the session presented a recent work in the fields of cloud network infrastructure design, mobile ad hoc networks, ontology-based data sharing, simulation by using agent-based models, crowd-sourced experimental platform and dynamic software adapters.

2) Regular Sessions

As mentioned before, the papers accepted to the regular program of the conference were classified in several relevant topics, after evaluation.

The session on Software Development discussed about bio-inspired optimization algorithms, REST-based android applications, social data mining and analyses of RDF data stores performance, graph layout and data mining algorithms.

Different technology application issues in the fields of energy, environment and sustainable development were discussed in a dedicated session. Technologies in the specific industries were discussed, namely, mining and power supply. Also, the problems of environmental assessment and monitoring and optimization of the integrated energy dispatch were addressed.

The session on Simulation and Optimization mostly dealt with the relevant problems in transport, addressing in specific railway line simulation, wagon accumulation time and tire design optimization.

The session on Information Systems addressed the topics of service network monitoring, business improvement, information security awareness. Some specific applications in the areas of research management, contract management and financial management were also presented.

The last but not the least, the topic of Information Society gathered researchers dealing with cultural heritage preservation, healthcare knowledge innovation, telecommunication market analysis, ICT regulatory approaches, e-government projects' modeling and research process management.

3) Poster Sessions

Poster sessions at ICIST conferences have history of excellent and vibrant discussions. They typically present technical achievements or conceptual work which is not considered mature for oral presentations.

The poster papers at ICIST 2015 were addressing all above-mentioned topics with addition of the area of learning management systems. They are printed in the Volume 2 of this book.

B. Invited Keynotes

The scientific program was also supported by the exciting keynotes from the distinguished speakers.

Dr. Richard Mark Soley, a Chairman and Chief Executive Officer of OMG ®, Executive Director of the Cloud Standards Customer Council, and Executive Director of the Industrial Internet Consortium, presented the opportunities, disruptions and standards of so-called Industrial Internet. This new paradigm was proposed in response to the anticipated "major disruptions in transportation, financial management, medical devices and other markets as Internet thinking moves into the industrial domain". Dr. Soley presented The Industrial Internet Consortium (IIC), which aims at bringing "together different players to build and manage test-beds for IoT-enabling industrial systems, in order to identify new products & services, as well as priorities & requirements for standards to support adoption of these ideas".

Orri Erling, a program manager for the Virtuoso Hybrid RDBMS, at OpenLink Software, presented recent trends in data management and effect of these trends to the development of Virtuoso DBMS. He sketched a vision of re-convergence of Relational database, RDF store and computational platform and illustrated this vision with the use cases in life sciences, semantic search and geospatial information management.

Vladimir Weinstein has been working in the internationalization and localization infrastructure team in Google since 2006. In his keynote, he discussed about managing engineers in Google, in specific: technical versus managerial path, identifying talent and mentoring, helping engineers build their careers, and many roles of an engineering manager.

III. ACKNOWLEDGEMENT

The editors wish to express a sincere gratitude to all members of the International Program Committee and external reviewers, who provided a great contribution to the scientific programme, by sending the detailed and timely reviews on this year ICIST's submissions.

The editors are grateful to the organizing committee of YUINFO conference for providing full logistics and all other kinds of support in setup of exciting scientific and social program of ICIST 2015.

Failure-Correction Simulation Tool Applied for Skull Prosthesis Modelling

Marcelo Rudek*, Gustavo Campana Mendes*, Andreas Jahnen**
* Pontifical Catholic University of Parana – PUCPR / PPGEPS, Curitiba, Brazil
** Luxembourg Institute of Science and Technology – LIST, Luxembourg
marcelo.rudek@pucpr.br, gustavo.campana@hotmail.com.br, andreas.jahnen@list.lu

Abstract— **The work presents a proposal of a computational tool applied to simulate synthetically built failures over skull bone images for use in prosthesis modelling studies. The main goal is to produce failures from known geometries to evaluate different filling processes used in bone repairing. The method is applied to the modelling of 3D prosthesis from computed tomography scans. A prototype software was developed as a plugin in Java based image processor (ImageJ). The key advantage is that testing failures can be built in a public domain image. Also the respective virtual prosthesis can be fully modeled and evaluated during study or surgical training. By using the lateral symmetry of skull, we applied a mirroring method as a study case to demonstrate the process, from the simulation of a failure up to evaluation of the solution.**

I. INTRODUCTION

Recent advances in 3D medical image processing have improved capabilities in human body visualization, providing important resources for diagnosis and treatment of diseases. These images obtained through CT (Computed Tomography) also allow us to understand the morphological characteristics of bones and its complex structures [1], [2]. Beyond visualization, the possibility to build or repair missing regions in bone structures through rapid prototyping devices and procedures are also in focus.

Actual manufacture devices, such as 3D printers, are important tools in the prototyping of complexes shapes, needed in anatomic prosthesis manufacturing, for example. An automated manufacturing procedure brings more accurate results when compared to hand crafted ones [3], [9], [12] as it is not dependent of the doctor's experience and ability, but only from data extracted straight from computed tomography images.

The integration between image analysis and prototyping has been explored by [3], [4] as part of an automated process of geometry modeling. In this way it is possible to generate the CAD profiles automatically for a machining process. However, the geometrical complexity of the individual human bones increases the challenge in modelling. The computational environment to medical image processing, called ImageJ [5], [6] provides the mechanisms to 3D modelling, visualization and also exporting data to manufacturing machines. The necessary data to build the geometric model of bones can be taken from tomographic scans extracted out of the DICOM (Digital Imaging and Communications in Medicine) file, according to the standardized protocol for computed tomography (CT) exams described in [7].

In a real case of repairing missing areas (holes) in the skull, the image information of failure's region is absent. Different strategies might be applied to generate a model that fills the open region [10-14].

Depending on the deformity location, e.g. in frontal or lateral position we have respectively the non-symmetrical and symmetrical cases. In symmetrical case, the intact bone of one side can be mirrored in order to repair a defective area on the opposite side. But in non-symmetric case, i.e. in frontal or rear position of skull we don't have information to be mirrored and another strategy must be used. Some filling algorithms are proposed in literature passing through all individual CT slices, as the Ellipse Adjustment Algorithm (EAA) proposed by [3] to fill a non-symmetric frontal failure. In this case the missing surface is reconstructed from arcs extracted from ellipses adjusted on skull's curvature. Despite the fact that this technique seems to be promising, the generated curvature do not always adjust correctly to the real skull shape.

A similar error condition also occurs in the case of symmetry by the mirroring strategy presented by [4], mainly in junction areas between skull and prosthesis. We address this condition further in the text. The fact is that still are some unresolved points about quality of automatic adjust of curves on CT slices and its respective strategies.

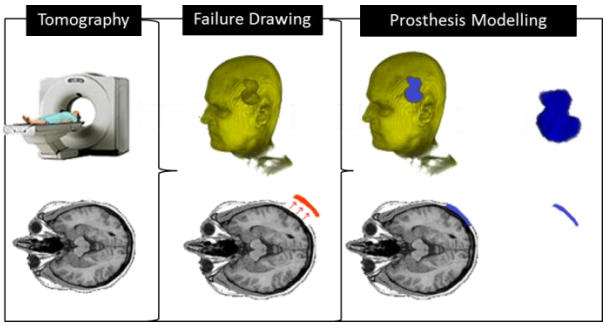
As a pre-processing step before manufacturing, we are proposing a computational tool to simulate failures in skull image with different geometries and also offer the capability to fix it. Thus it is possible to evaluate the prosthesis modelling methods by testing different conditions of failures. With this proposal, we were able to compare the generated solution with the real information previously extracted from original image. In this way we will have better conditions (parameters) to evaluate which method is the most suitable to be applied in a real case with a similar condition.

In order to illustrate the method, it is presented as a study case of the repair of a synthetically created failure in the lateral of skull. In this case we are adopting a mirroring process by symmetry. In addition, the skull and the model of prosthesis were prototyped by a 3D printer, thus the matching between printed skull and printed prosthesis can be compared with its real shape and dimensions.

The goal is to produce a simulation environment for prosthesis modelling without a dependence of real patient images. Thus we have a research tool for studies, training and procedures' planning in the context of prosthesis modeling, both in the medical and engineering point of view.

II. PROPOSED METHOD

The interest here is to create a manner to test several different modelling algorithms, and evaluate them by comparing with original known information (original existing bone). The proposed method is composed of three main steps as shown in fig.1. First, tomographic scans are needed to obtain the bone shape, i.e. the curvature in each section of the skull. Secondly, a failure area can be drawn to define a region of interest (ROI), thus this region can be cut and extracted out from the skull image, creating a hole with the desired shape. And finally in the third step, a specific method can be used to fill that region, i.e. the prosthesis modelling.



(a) (b) (c)

Figure 1. The three sequential steps of proposed method. (a) A tomography scan; (b) Failure drawing over it by cutting a piece of bone and (c) the respective prosthesis modelling to complete the gap.

The cuts are applied from the drawing, slice by slice, and then their surfaces are reconstructed in 3D. This extracted piece element can be used for further comparison with the new the prosthesis modeled.

The software was developed as a plugin to be installed in the ImageJ [5]. The ImageJ is a free open source image processing software environment, and the new plugin functionalities can be combined with its image processing facilities. The new plugin was developed to cut the ROI slice by slice, extract the 3D surface and apply the mirroring algorithm to generate the new piece.

The new piece (prosthesis) can be compared with the original bone through the GeoMagic® Studio software [8], which allows the measurement of errors between both pieces, that are evaluated and visualized in 3D surfaces. If necessary the ImageJ can generate the STL profile format for 3D printing or other manufacturing processes.

The three steps of the method are presented in following sections.

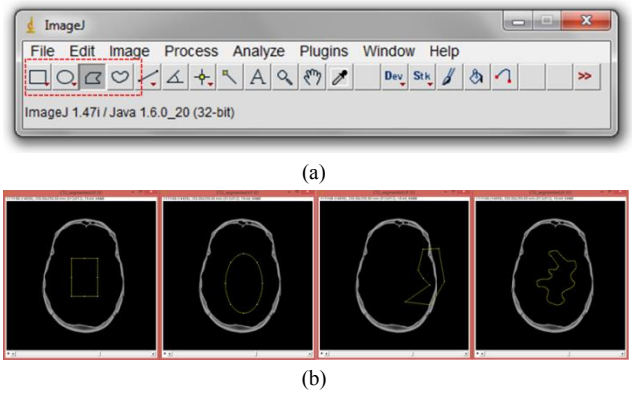
A. CT Slices Segmentation

The CT Slices containing the skull image extracted from a DICOM file [7]. In this investigation we are using a public domain image from [15] containing the tomography of a human head, but any other real image can be used in the same way.

The bone segmentation, from different biological materials e.g. brain or fluids, is performed by ImageJ toolbox on the set of selected CT slices permitting to define the edges of skull in each one of them, composing a black and white image of the contour bone.

B. Failure Drawing

The second step on process is to define the ROI (Region of Interest) to be applied to skull. It is defined as the cutting position of bone edge. A region is taken out from image by selecting a polygonal (or not) area through the selection cutting tool from ImageJ menu as highlighted in fig. 2.a. In fig. 2.b. there are four examples of cutting areas on a CT slice in top view. The cutting area is defined on the first slice selected and then forward propagated to all other necessary slices in z direction (entering in image). The cutting tool can be used in any plan, i.e. transversal, sagittal and frontal plans.



(a)

(b)

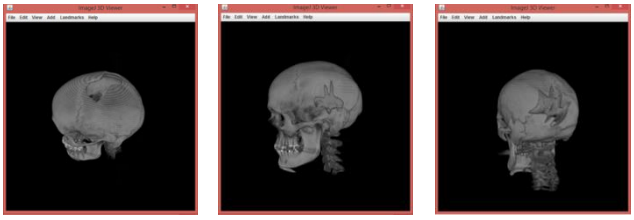
Figure 2. (a) Selection tool. (b) Samples of cutting application.

Thus it is possible to create a defective area with randomized geometry on each CT slice and after rebuild the 3D image of skull and of the removed piece. Let I_o^k as original m x n image matrix extracted from DICOM whose k value is the slice number, we have that,

$$I_o^k = I_s^k + I_p^k \quad (1)$$

where I_s^k is the matrix with resting of pixels kept in original skull image and I_p^k is the matrix with the removed piece for each testing slice k.

Each slice k can be cut as wished and thus after 3D reconstruction a hole is built in the skull, forming the image I_s^k. The new image I_s^k is the one used to test the filling algorithms. Fig. 3 shows three examples of synthetically created defects. In 3.a is the top view in transversal plan; 3.b is an example showing the lateral side of the skull with defect created in sagittal plan, and 3.c is a defected created in the rear of the skull in frontal plan.

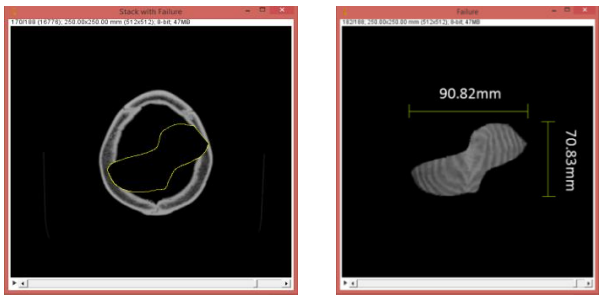


(a) (b) (c)

Figure 3. Samples of failures syntactically created in the plans transversal, sagittal and frontal.

The image I_p^k contains the pixels' matrix of the extracted piece from the original image. After the 3D

reconstruction from k slices of I_p^k we have a surface of the extracted piece as presented in fig. 4. This piece will be used as a comparison element with the new prosthesis model. The dimensions of the extracted piece can be seeing in *mm* as shown in fig. 4.



(a) (b)

Figure 4. Piece extracted from original skull. (a) CT slice near of skull top. (b) The surface of extracted piece.

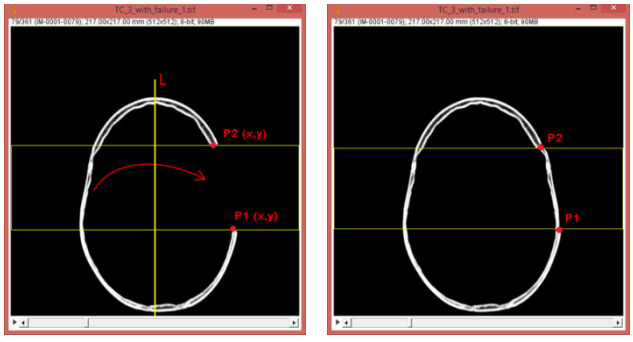
Fig. 4.a shows a random shape cutting area applied to the first slice. A complete piece was extracted from eight slices from top down direction as shown in fig. 4.b. It causes a synthetic failure in the 3D visualization of the skull (as presented previously in fig. 3.a) given by I_s^k matrix. From now, the testing image is I_s^k and the next step is to find the prosthesis matrix ϕ, where

$$\phi = \sum_{k}^{n} I_p^{*k} \quad , k = 1, \dots, n \tag{2}$$

In (2) the sum operator signifies the superposing of k slices into matrix ϕ, i.e. the size of ϕ is $m \times n \times k$, with $k = 1,\dots, n$ sequential slices obtained from the new matrix I_p^{*k}. The new matrix I_p^{*k} will be obtained by symmetry, with skull edge pixels mirrored from the opposite side of failure (considering a lateral failure). Then, ϕ is obtained from all superposed k mirrored images.

C. Prosthesis Modeling

The third step of the method is to fill the open region by building the image matrix I_p^{*k}. Among several methods applied to prosthesis modeling, we are using a mirroring method to copy the known region symmetrically from one side (closed) to opposite side (open) as presented in fig. 5. We are considering only failures in lateral position of skull as divided by the symmetry line L.



(a) (b)

Figure 5. Mirroring by symmetry on the gap position.

In this case, the assumption is that if a region is unknown, then the more approximated corresponding values to fill it can be obtained by symmetry. In fig. 5 the P_1 and P_2 values are the known coordinates from interruption points in the edge of right side. These points are found by the method proposed in [3]. Using the same distances from x and y values, we obtain the respective coordinates of limits from left side that contains the complete edge information. Thus the region between points P_1 and P_2 are now completed with mirrored piece as method of [4].

III. Evaluation Of Method

In order to evaluate the method we build a synthetic failure as a single polygon as shown in fig. 6. This shape permits to see better the junction points and to reduce the complexity to prototype (3D print) instead of another complex geometry.

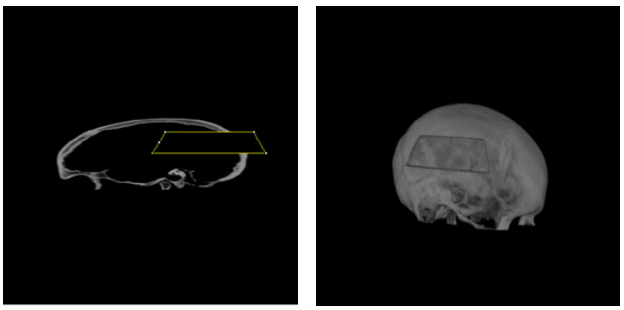
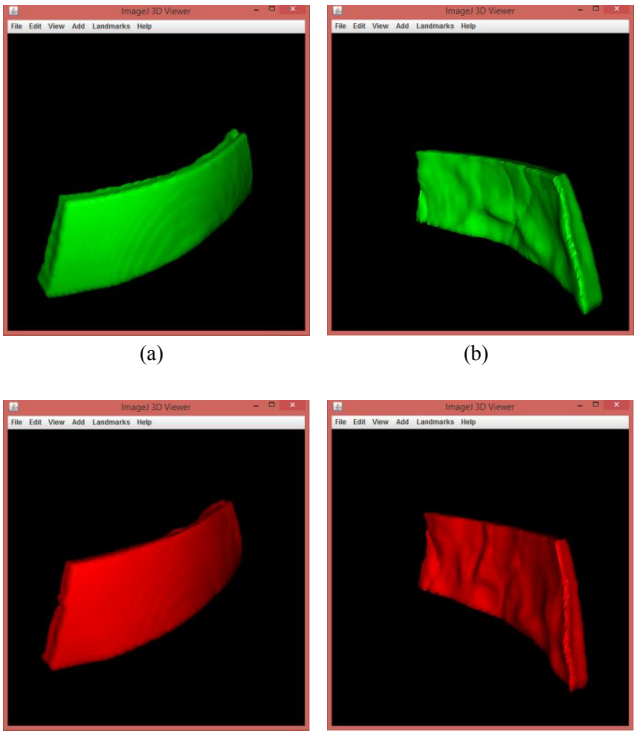


Figure 6. (a) Cut in edge of a CT slice. (b) New 3D surface with created failure.

The process is repeated for all open slices and a surface is obtained after all slices have been processed as in fig. 7.



(a) (b)

(c) (d)

Figure 7. 3D views of pieces: original (green) and modeled piece (red).

The Fig. 7.a and 7.b are the external and internal view of original bone data I_p^k; 7.c and 7.d are the external and internal view of modeled prosthesis image matrix ϕ by mirroring.

By a visual sense, the shape of I_p^k (original bone) and ϕ (modeled prosthesis) are similar, but by measurement it is not totally true. The displacement between original bone and modeled prosthesis can be evaluated by GeoMagic Studio software in 3D and the result is shown in fig. 8.

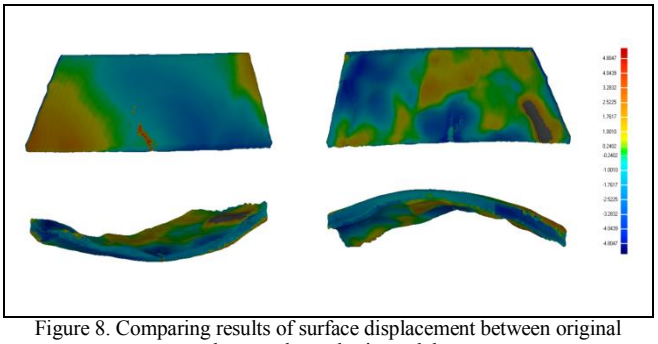
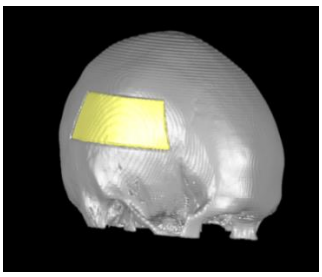
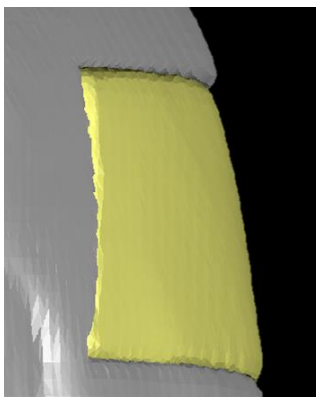


Figure 8. Comparing results of surface displacement between original bone and prosthesis model.

The change in colors in figure 8 shows the difference in *mm* between I_p^k (original bone of real surface in ROI) and its respective matrix ϕ (model of prosthesis) created from mirroring. The maximum error's value is about ~ 4.8 *mm* in the inner side due to more roughness in its surface, and maximum of ~ 3.2 *mm* in the outer side. The prosthesis ϕ applied on failure hole can be visualized in figure 9.



(a)



(b)

Figure 9. ROI filled by prosthesis. (a) filling the skull gap, and (b) difference in junctions between skull and ϕ model.

In fig. 9.a the original skull was filled by modeled prosthesis ϕ. And, the fig. 9.b is the prosthesis matching and its respective visible error.

In additional we printed a created 3D model. The ImageJ software is able to export data to machining by STL formatted file. A 3D printer was used to build the real piece based on modeled prosthesis ϕ. Fig. 10 presents the results.

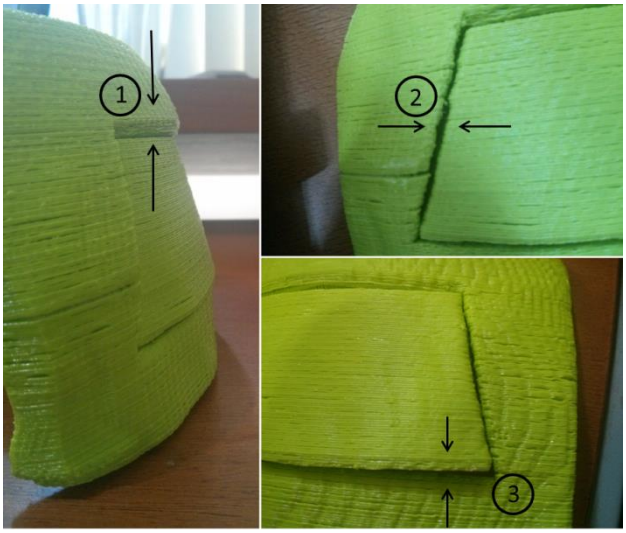


Figure 10. The 3D print of modeled prosthesis.

The fig. 10 presents three labeled regions in different views of modeled prosthesis. There are differences in relationship of original skull bone. In the pictures labeled as '1' and '2' the piece is entering into original skull. In highlighted picture '3' the corner of the piece is outside of the original skull bone alignment. It is possible to see that there are a significantly difference between the original bone piece I_p^k and the synthetically build ϕ. Thus, despite symmetry of human body, the mirroring does not bring any guaranties of a correct adjustment of the prosthesis piece.

IV. CONCLUSION

The research presented the proposal of a simulation software tool in order to build synthetic failures in skull border and its respective correction procedure by mirroring. A handmade synthetic failure can be drawn on the skull, performed slice by slice in a selected direction as, e.g. frontal, sagittal and transversal plans. A filling strategy by mirroring was applied to produce the respective prosthesis model. The new modeled piece could be compared with the original extracted piece of bone inside ROI. By 3D comparing between original bone and model, we are able to see the measurements errors in the modeled piece. The surface of the prosthesis was 3D printed and the matching point's problems become visible as equal indicated in the geometric model. Despite 3D differences in measures, the tool seems to be a promising tool to studying and modelling prototyping. The next step is embedding new procedures of prosthesis modelling to both symmetric and no-symmetric cases.

ACKNOWLEDGMENT

The authors would like to thanks the Pontifical Catholic University of Parana – PUCPR in Brazil and the Luxembourg Institute of Science and Technology – LIST by financial and technical support.

REFERENCES

[1] M. Rudek, O. Canciglieri Jr., A. Jahnen and G. L. Bichinho, "CT Slice Retrieval by Shape Ellipses Descriptors for Skull Repairing". IEEE International Conference on Image Processing ICIP 2013, pp. 761-764, 2013.

[2] H. Li, Z. Xie, S. Ruan and H. Wang, "The Measurement and Analyses of Symmetry Characteristic of Human Skull Based on CT images". Biomechanics and Vehicle Safety Engineering Centre, Tianjin University of Science and Technology, Tianjin, China, vol. 26, n. 1, pp. 34-37, 2009.

[3] T. Greboge, M. Rudek, A. Jahnen and O. Canciglieri Jr., "Improved Engineering Design Strategy Applied to Prosthesis Modelling". *Product Service Engineering in a Dynamic World*. 1ed. Amsterdam: IOS Press, pp. 60-71. 2013.

[4] G. C. Mendes, "Proposta de um Método para Simulação de Defeitos em Imagens Tomográficas Aplicado na Modelagem Geométrica e Confecção de Próteses Anatômicas de Crânio". (in Portuguese), Scientific Research Seminar, PUCPR, 2014.

[5] ImageJ, Open Source Software Project, Available online at <http://imagej.net/ Development> Accessed in Oct. 2014.

[6] J. Schindelin, et al., "Fiji: an open-source platform for biological-image analysis", Nature Methods 9(7): pp. 676-682, 2013.

[7] DICOM, "Digital Imaging and Communications in Medicine Part 5: Data Structures and Encoding", Available online at < www.medical.nema.org/dicom>, Accessed Nov. 2014.

[8] 3D Systems - Geomagic Solutions, Idaho Visualization Laboratory, USA. Available online at < http://geomagic.com/en> Accessed Nov 2014.

[9] B. J. Kim et al., "Customized Cranioplasty Implants Using Three-Dimensional Printers and Polymethyl-Methacrylate Casting". *J. Korean Neurosurg Soc.*, Seoul, 2012.

[10] M. H. Mulroy et al., "Evaluation of pediatric skull fracture imaging techniques". *Forensic Science International*, v. 214, n. 1-3, 2012.

[11] B. D. S. Pinto, A. D. C. Ribeiro, A. Sousa, "Reengenharia de Sistema Produtivo Integrado para Fins Educacionais: Conceitos Gerais CAD/CAM/CAE/CIM". Faculdade de Engenharia da Universidade do Porto (FEUP). Porto, pp. 25. 2005.

[12] F. I. Saldarriaga, et al, "Design and Manufacturing of a Custom Skull Implant". *American J. of Engineering and Applied Sciences*, v. 4, n. 1, 2011.

[13] L. H. Stieglitz, et al. "Intraoperative fabrication of patient-specific molded implants for skull reconstruction: single-centre experience of 28 cases". *Acta Neurocirurgica*, n. 4, pp.156, 2014.

[14] F. Xiao et al. "Estimating postoperative skull defect volume from CT images using the ABC method". *Clinical Neurology and Neurosurgery*, v. 114, n. 3, 2012.

[15] PCIR Researchers, Patient Contributed Image Repository, Available online at <http://www.pcir.org/researchers/ downloads_available.html> Accessed in Oct. 2014.

Implementation of the Smartphone Based Biofeedback Application

Anton Kos and Anton Umek
Faculty of Electrical Engineering, University of Ljubljana, Slovenia
anton.kos@fe.uni-lj.si, anton.umek@fe.uni-lj.si

***Abstract*—Biofeedback applications can prove useful in many areas of human activity. For example, with a system for motion tracking and the use of biofeedback one can be guided to learn the proper movement or informed of its improper execution. This could be especially useful in sports and rehabilitation. In a biofeedback system the feedback information is communicated back to the user, preferably in real time, through one of the human senses: sight, touch, and hearing. The implemented movement tracking biofeedback system requires: inertial sensor(s), a processing device and a biofeedback device. Today smartphones are readily available and more often than not include inertial sensors. The development of a smartphone based biofeedback application is therefore a logical step. The developed application tracks head movements during the execution of the golf swing. Improper head movements are detected and communicated back to the user in real time in a form of audio signals. The user acts on received biofeedback signals trying to correct the improper movement. Extensive measurements and tests were performed to confirm the correct and accurate operation of the application. The usefulness and advantages of real-time biofeedback during the golf swing execution were identified. Golf players with excessive and improper head movements considerably improved their performance after the inclusion of the biofeedback. We believe that such biofeedback systems are applicable to similar examples in sport, fitness, healthcare, and other areas of activity.**

I. INTRODUCTION

Biofeedback can be very useful in sports, rehabilitation, and other human activities. In sports and rehabilitation one can learn the proper movements with the help of biofeedback. For instance, the biofeedback can be used to direct or instruct the user to properly perform the movement or alert the user in case of improper movement execution. For the movement analysis a motion tracking system is necessary. A motion tracking system can be based upon different technologies; some examples being: video recordings, optical tracking system, and a system with inertial sensors.

Biofeedback can be applied in real time, during the movement execution, or out of the real time, after the movement has been completed. The example of the latter is the inspection of the recorded video of the performed movement or the post processing of recorded sensor signals as in [3]. The interruptions caused by the non-real-time biofeedback slow down the movement learning process. A better choice is the use of the real-time biofeedback where the feedback information is communicated to the user during the movement execution.

Several channels corresponding to human senses can be used for the biofeedback: visual, tactile, and auditory. The visual channel holds the most information, but it requires high cognitive load and it has relatively high reactive time. The tactile channel has lower reactive time, but it can be irritating or even painful for the user and relatively difficult to implement. The auditory channel has low cognitive load, low reaction time and it is relatively easy to implement.

Nowadays smartphones are readily available, wide-spread technology and in many countries the penetration of smartphones has exceeded 50% in 2014 [1]. The majority of smartphones include inertial sensors that are used for a number of mostly simple applications. We chose smartphone inertial sensors to build a biofeedback system that tracks user movements, detect deviations from the expected positions, and communicate the possible movement errors in real time to the users through the auditory channel.

The use of smartphone inertial sensors for motion tracking is limited due to their relatively low quality which is expressed through the imprecision of sensor signals [4]-[7]. While the position deviation in short-time tracking may be small enough to be neglected, the deviation for long-time tracking is most often too large to be disregarded. One of the motivations for this work is the determination of the usability of inexpensive inertial sensors integrated into today's smartphones for the implementation of a real-time biofeedback application.

The biofeedback application for head movement tracking in golf was developed. Tests on several groups of golf players were conducted. The test show that real-time biofeedback speeds up the correct movement learning process.

This paper is organized as follows. In section II we give the definition of a biofeedback system, explain its building blocks and its functionality. We continue with the demonstration of the biofeedback system operation, the explanation of its versions that are suitable for our application, and the discussion of the most appropriate user interfaces. Section III presents the biofeedback application, its demands, constraints, building blocks, configurations, and user interfaces. In section IV we present the results of application test and biofeedback experiments. We conclude with section V.

II. BIOFEEDBACK SYSTEM

In a *biofeedback* system a person has attached sensors for measuring body functions and parameters (*bio*). Sensors are connected to a computer or any other device

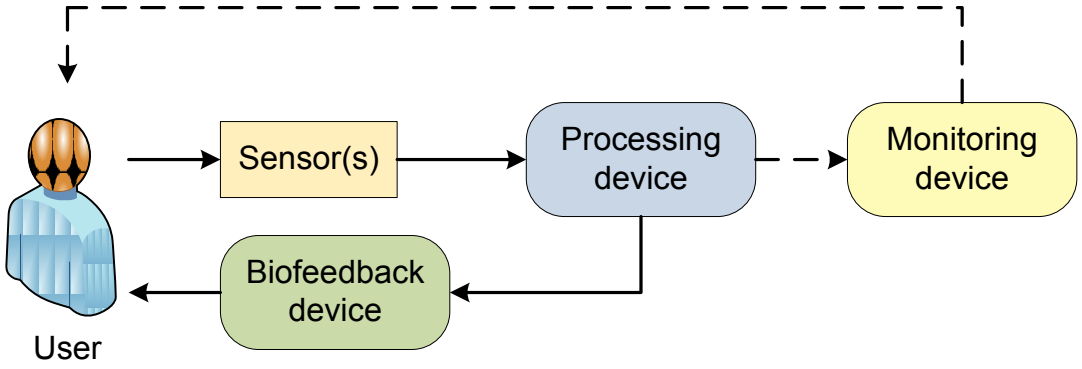


Figure 1. An example of a compact version of the biofeedback system. All system devices are at the user location, therefore the real time operation of the biofeedback loop is not a problem. Monitoring device is not mandatory; it is generally used for reviewing the results.

for analyzing data. Results are communicated back to the person (*feedback*) through one of the human senses (i.e. sight, hearing, touch) [2]. The person tries to act on the received information in order to change the body function or parameter in the desired way. The term biofeedback is frequently described in the connection with physiological processes; in this paper we use it in the connection with the body activity in the sense of physical movement. We use inertial sensors to detect the body movement, processing devices to analyze sensor data, and audio signals for the (bio)feedback.

In general, our biofeedback system consists of several different autonomous devices which are interconnected through wireless communication channels. The primary system tasks are: real-time movement tracking and, in connection with the later, real-time biofeedback.

A. *System building blocks*

In general a biofeedback system requires: inertial sensor(s), a processing device, biofeedback device(s), a monitoring device (optional), and in the majority of versions also wireless communication channel(s).

Inertial sensors (accelerometers and gyroscopes) are the essential part of the system, which is designed to work with one or multiple autonomous sensor devices. *Processing device* is the core of the system. It receives inertial sensor signals, analyzes them in real time, and when needed, it generates and sends feedback signals to biofeedback devices in real time. *Biofeedback device* employs our senses to communicate feedback information to the user. Commonly used senses are: hearing, sight, and touch. In our system we use audio feedback devices such as loudspeakers and headphones. *Monitoring device* is optional. It is used to monitor sensor signals and analysis results. The monitoring can take place in real time or as a post experimental results analysis. *Communication channels* enable the communication between the system devices. While wireless communication technologies are most commonly used, wired technologies can also be practical; for instance to send a feedback signal from the body-attached processing device (i.e. smartphone) to the nearby headphones.

B. *System operation*

The generalized operation of the biofeedback system, composed of the above described building blocks, is shown in Fig. 1. Sensor(s) continuously send inertial signals to a processing device for analysis. The activity of the biofeedback device and thus the activity of the feedback loop depend mostly on the analysis results. The basic loop operation modes are:

- *standby*: the processing device is continuously analyzing sensor signals, but a biofeedback signal is not generated; the feedback loop remains open,
- *user guidance*: the biofeedback signal is constantly generated to guide the user to perform a certain movement or to assume a certain position; the loop is constantly closed and the user is expected to react to the biofeedback signal in real time,
- *error detection*: the biofeedback signal is generated only when the system detects an improper movement; the loop is closed for shorter time periods and the user is less likely to be able to react in real time.

C. *System versions*

Two basic versions of the presented biofeedback system are distinguished: the *compact* version, where all the system devices are attached to the user, and the *distributed* version, where some of the system devices are at a remote location, away from the user. For example, in the distributed version the processing device and the monitoring device from Fig. 1 are at the remote location, sensor(s) and the biofeedback device are at the user. All devices in both locations are typically interconnected through wireless communication channels.

In the compact version of the system the biofeedback loop is located at the user; hence the real-time operation is not compromised due to the communication channel latencies. In the distributed version the biofeedback loop is distributed between the user and the remote location. To assure the real-time operation of the loop, the distributed version of the system must use low-latency communication channels and is essentially bound to the areas of limited extent.

Smartphones can be used in both biofeedback system versions. In the compact version the smartphone is attached to the body of the user and provides functions of all system devices: integrated inertial sensor(s), processing power, monitoring device (screen), and biofeedback device (speakers). In the distributed version the phone's wireless communication interfaces are used to connect to the remote devices.

D. *User interface(s)*

Different versions of the system commonly have different user interfaces. For instance, in the distributed versions of the system the interface is divided between the user and the remote location. Since users are performing movements, they generally do not have the possibility to control the operation of the system by using standard interfaces; therefore the operation of the system is best driven by user gestures. Gestures are defined by user's body movements and detected through their characteristic inertial sensors responses. Each biofeedback application has its unique set of gestures that are adapted to its movement patterns. Details about the gesture user interface developed for our application are found later in the text.

III. REAL-TIME BIOFEEDBACK APPLICATION

Based on the biofeedback system presented above the real-time biofeedback application for golf was developed and implemented. The main functionality of the application is real-time golf swing analysis with real-time audio biofeedback depending on golfer's head movements.

A. *Basic application idea*

Our basic application idea is to monitor the player's head movements during the golf swing. The idea is grounded on the hypothesis, that wrong movement patterns in the golf swing many times lead to distinct unwanted head movements. Consequently, head movements are very often the indicator of the incorrect golf swing execution. Our hypothesis is backed-up by observing the world's best golf players and their advices [8] and [9]. The majority of them are keeping their heads practically still; hence our biofeedback idea is: "golfers head should stay still during the swing execution". Therefore, the application is developed to detect excessive incorrect head movements and communicate them to the user in real time using the audio feedback.

B. *Application demands and constraints*

The most important demands are the real-time operation of the biofeedback loop and the implementation of a non-distracting user interface.

One of the constraints faced by the system is the inaccuracy of sensor signals, that primarily limit the time frame of the analysis and puts the movement detection repeatability under question. Typical duration of the golf swing is around 2 seconds [10]. The authors in [11] state that by using the uncompensated sensor signals of the iPhone 4 smartphone the derived angle and position errors are 1 deg/s and 19 cm after the first second respectively.

Apart from the above mentioned demands and constraints that must be met by the application, there are functionalities that are not binding, but may contribute greatly to the overall usability and user experience of the application. Some, but by no means the only such functionalities, are the ease of use (connected to the user interface and the application design) and the possibility of post processing of the data collected by recording of swings.

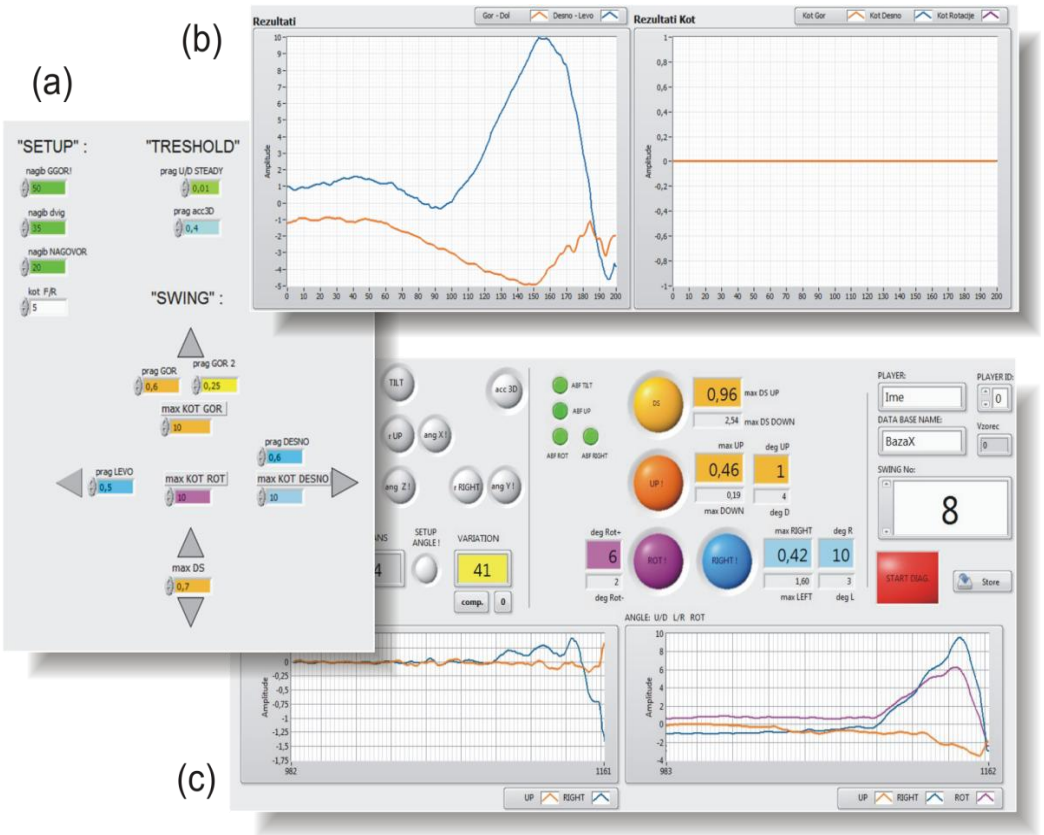


Figure 2. A graphical part of the user interface with keyboard and mouse as input devices. The application is controlled through several windows or tabs such as: (a) settings, (b) swing history, (c) real-time signal monitoring and system states.

C. *Application building blocks and configuration*

For practicality, we first developed a distributed version of the biofeedback system (see section II.C). In this version the biofeedback loop is distributed between the user and the remote location. Inertial sensors and the biofeedback device are attached to the user's head, the processing and monitoring devices are at the remote location. The devices on both locations are connected through low-latency wireless channels.

The application uses the inertial sensors integrated into the iPhone 4 smartphone, which is attached to the head of the golf player. With the appropriate attachment of the smartphone to the head of the golf player we achieve very good repeatability of detection of different 3D head movement measured from the static start position at the swing setup.

A laptop is used for the processing device. The laptop must be on the location which is in the range of a local wireless network used for the transfer of sensor data from the smartphone to the laptop. The application is designed in the LabVIEW™ development environment and is able to run on any compatible MS Windows operating system.

Headphones are used for the audio biofeedback device. They are wirelessly connected to the laptop. The audio feedback signal is generated by the laptop and sent to the headphones over the RF-ISM channel.

D. *User interface*

Our application implements a hybrid user interface comprised of a graphic user interface for controlling the application on the laptop and a gesture user interface for controlling the application during the execution of the swing. Both interfaces work simultaneously.

The graphical user interface runs on the laptop and allows the complete access and control of the application and its functionality at all times. Fig. 2 shows the graphical user interface and some application windows.

The gesture driven user interface greatly eases the application usage; it allows the user to fully concentrate on the swing itself, and not on controlling the operation of the application. The application and its gesture user interface (GeUI) are able to detect different golf shot stages and swing phases. For instance; gesture user interface helps users to take the correct swing setup position.

The indispensible parts of the user interface are the characteristic audio signals. They give the user the information about the application states and transitions between them, they inform the user that the application is ready for the swing; they signal errors to the user, etc.

E. *Real-time signal analysis*

During the golf swing training, the user should be focused primarily on the swing execution and the possible acoustic feedback signals. To alleviate the user's interaction with the application at the times of training the GeUI was developed. The application also includes a window with abundant data and information about the swing, which are simultaneously shown in numerous diagrams, error indicators, peak value detectors as shown in Fig. 2. The user can review them after the swing. If needed, the user can modify the application parameters between each swing.

IV. RESULTS OF APPLICATION TEST AND BIOFEEDBACK EXPERIMENT

We present the most significant results of numerous measurements and experiments aimed at:

- testing the correct and precise application operation,
- usability of real-time biofeedback for the correction of errors during the performance of a golf swing.

A. *Application test*

During the application test we focused primarily on: (a) correct detection of application states and transitions between them, and (b) precision of sensor signals and the derived analysis results.

The application test stage includes the recording of golf swings performed by a professional golf player with a very consistent swing execution in terms of movement repeatability. We have recorded all of the performed swings without any selection. The results are shown in Fig. 3 and they confirm the following:

- The correct detection of application states and the transitions between them. This is evident from the time alignment of the signal peaks.
- The sufficient precision of sensor signals. This is evident from almost perfect alignment of curves representing the swings and their identical shape. We can observe that all the curves begin (takeaway swing phase) precisely at rotation angle zero and deviate for less than a few degrees throughout the swing.

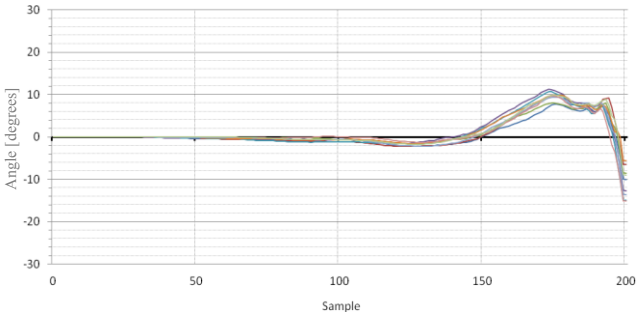


Figure 3. Series of swings performed by a professional golf player. Curves shows head rotation angle in degrees in the Left/Right (+/-) direction for the last 200 samples of the swing. The sample frequency is 60 Hz, head rotation to the left yields positive angle values. We observe that the head movement is very consistent for all executed swings

From the results in Fig 3. it can be concluded that the application works correctly and that the precision of the sensor signals is high enough for the analysis of the required quality. The acquired records of the professional golf player also serve as the source and model for setting the threshold values for triggering and state signals used by the application.

B. *Biofeedback experiment*

In the biofeedback experiment we close the biofeedback loop by activating the biofeedback signals. Head movements exceeding defined thresholds are treated as errors and are communicated to the user in real time, during the entire execution of the swing. The biofeedback signal in the application is a binary acoustic signal (beep).

When users are aware of the cause of their swing errors, they can easily understand the biofeedback signals and consequently act on them. The application allows that the biofeedback is triggered at different combinations of available head movement signals. To prove this concept we conducted a series of experiments that would show that using the real-time audio biofeedback could help golf players with inconsistent swing correct their unwanted excessive head movements and hence improve their swing consistency and performance.

Fig 4. shows the results for the beginner player that excessively moves the head during the execution of the swing. The curves represent the average head rotation speed in the left/right direction calculated from the series of twenty swings. The dashed line represents the average for swings without the (audio) biofeedback, meaning that the player performed the swings without the notifications about the detected errors. The head rotation speed error threshold was set to the signal value ±0.15 rad/s. It can be noticed that the player considerably exceeds the set thresholds. After switching the biofeedback signals on, the player performed another series of twenty swings. Their average is represented by the solid line. It is evident that the biofeedback considerably helped the player. The left/right head movement has been almost eliminated.

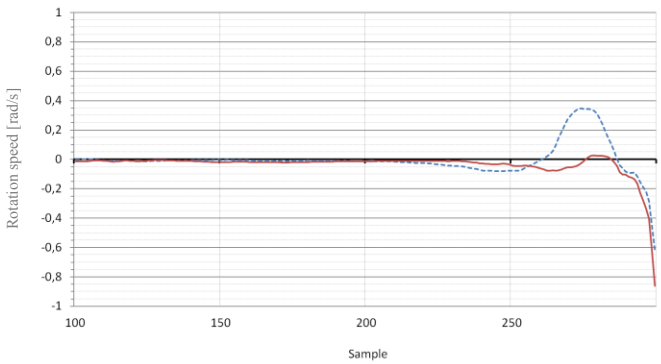


Figure 4. The benefit of biofeedback for a beginner golf player. The graph shows the comparison of player's head rotation speed [rad/s] in the Left/Right (+/-) direction without the biofeedback (dashed line) and with the biofeedback (solid line). Curves represent the averages calculated from twenty swings performed without the biofeedback and twenty swings performed with the biofeedback.

We must emphasize that the above are only the preliminary results of biofeedback experiments conducted on a limited population. The results were primarily gathered for the purpose of application test and the test of the biofeedback loop operation. More extensive experiments are required to prove the usability of such systems in sports, rehabilitation and other possible areas of use. What surprised us is the fact that the learning process was even faster than anticipated. At this point we do not wish to speculate about the movement learning process or corrections mechanisms with the aid of biofeedback. To prove the usefulness and benefits of biofeedback more extensive experiments and analysis are required, which will incorporate both: the objective measures (signals) and subjective measures (user opinion).

V. CONCLUSION

This work shows that the existing smartphones with their built-in inertial sensors can be used for short-time movement analysis in real-time biofeedback systems.

For the implementation of the test biofeedback system we chose golf. We developed the application that uses audio feedback for the correction of golf swing in real time. We also developed a hands free gesture user interface for the user. We are convinced that the gesture user interface is an indispensable element of the biofeedback system, because it alleviates user's interaction with the system.

Field experiments show that our application is an efficient tool for the correction of the head movement errors, especially for beginner players having problems with swing consistency.

The next challenge and the logical step forward is the implementation of the compact version of the biofeedback system. The compact version runs entirely on the smartphone without the need for a separate external processing device and without the need of wireless communication channels.

While we have tested our biofeedback application on the golf swing example, we believe that such biofeedback systems are applicable to similar examples in sport, fitness, healthcare, and other areas.

REFERENCES

[1] Global smartphone-penetration 2014. Available online: https://ondeviceresearch.com/blog/global-smartphone-penetration-2014

[2] Oonagh M Giggins, Ulrik McCarthy Persson, Brian Caulfield, Biofeedback in rehabilitation. Journal of NeuroEngineering and Rehabilitation, 2013

[3] Sara Stančin, Sašo Tomažič, "*Early improper motion detection in golf swings using wearable motion sensors: the first approach*", Sensors, 2013, vol. 13, no. 6, pages 7505-7521

[4] Anton Umek, Anton Kos, "*Usability of Smartphone Inertial Sensors for Confined Area Motion Tracking*", ICIST 2014 4th International Conference on Information Society Technology, pages 160-163, Kopaonik, Serbia, 2014

[5] "*Motion sensing in the iPhone 4: MEMS accelerometer*", Available online: http://www.memsjournal.com/2010/12/motion-sensing-in-the-iphone-4-mems-accelerometer.html

[6] ST Microelectronics, "*Everything about STMicroelectronics' 3-axis digital MEMS gyroscopes*", TA0343 Technical article. ST Microelectronics, July 2011

[7] Mohinder Grewal; Angus Andrews, "*How good is your gyro*". IEEE Control Systems Magazine, February 2010

[8] Tiger Woods, "*Maintain A Quiet Head*". *Available online:* http://www.golfdigest.com/golf-instruction/2009-10/tiger_woods_keep_quiet_head

[9] Bob Doyle, "*Experts Weigh In On Head Movement During The Golf Swing*", Available online: https://foreverbettergolf.com/articles/experts-weigh-in-on-head-movement-during-the-golf-swing/

[10] *"Golf Swing Timings*", Available online: http://www.smarthomepro.com.au/igadgets/92-golfsense-personal-3d-golf-sensor.html.

[11] Anton Umek, Sašo Tomažič, and Anton Kos, "*Autonomous Wearable Personal Training System with Real-Time Biofeedback and Gesture User Interface*", Proceedings of the 2014 International Conference on Identification, Information and Knowledge in the Internet of Things,pages 122-125, October 2014, Beijing, China

Application of Data Mining Algorithms for Detection of Masses on Digitalized Mammograms

Milos Radovic[1,2], Marina Djokovic[1], Aleksandar Peulic[1], Nenad Filipovic[1,2]
[1] Faculty of Engineering, University of Kragujevac, Serbia
[2] Research and Development Center for Bioengineering, BioIRC, Kragujevac, Serbia
mradovic@kg.ac.rs, marina.m.djokovic@gmail.com, aleksandar.peulic@kg.ac.rs, fica@kg.ac.rs

***Abstract*—This study presents the CAD (computer aided diagnosis) system for mass detection on digitized mammograms. The proposed system consists of four major steps: image preprocessing, image segmentation, feature extraction and classification. In the preprocessing step, contrast enhacement and 2-D median filtering has been performed. Segmentation step includes background and pectoral muscle removal, and detection of suspicious areas. In the following, feature extraction step, these suspicious areas are described with total 106 numerical attributes and afterward classified into normal or mass tissue by the use of seven different classifiers in classification step. By performing segmentation 92% of masses were correctly segmented with 4.14 false positives per image (FPpi). This result is improved in the classification phase where multilayer perceptron neural network (MLP) achieved 77.4% sensitivity and 0.49 false positive per image. We used images from the mini-MIAS database.**

I. INTRODUCTION

Breast cancer is the most common cancer among women in the world. Breast cancer screening with mammography has been shown to be effective for preventing breast cancer death [1]. Even for the well-trained radiologists reading mammograms is a very difficult task. An important development that may help to improve the performance in breast cancer screening is computer aided diagnosis (CAD).

A lot of research has been done in the field of computer aided detection of masses in digitalized mammograms. Cheng et al. [2] discussed different methods for mass detection and classification and compared their advantages and drawbacks. In this paper, methods of five major research areas: preprocessing, image segmentation, feature extraction and selection, mass detection/classification, and performance evaluation have been discussed in detail.

Nguyen et al. presented an automated CAD system for detection and classification of massive lesions in mammographic images [3]. In this work the authors performed ROI (region of interest) extraction by using edge-based mass detection algorithm, described each ROI with GLCM features and performed ROI classification by using artificial neural network. The authors tested proposed methodology on mini-MIAS database and achieved results including 3.47 FPpi, sensitivity of 85% and AUC (area under the ROC curve) 0.815 by using multilayer perceptron neural network.

In this paper we propose CAD system for mass detection on digitalized mammograms. Proposed system consists of four steps: (1) Image preprocessing, (2) Image segmentation (3) Feature Extraction and (4) Classification. In the preprocessing phase, contrast enhacement and 2-D median filtering has been performed. Segmentation phase includes background and pectoral muscle removal, and detection of suspicious areas. For detection of suspicious areas, simple regression function has been proposed as a threshold function. In order to maximize tumour detection sensitivity (number of correctly segmented masses) of preprocessing and segmentation phases optimization procedure has been performed. Feature extraction is important step in breast cancer detection process. In this phase we describe every suspicious region by numerical attributes calculated by using different methodologies. In this paper we calculate 11 statistical features, 20 GLCM (gray level co-occurrence matrix) features, 5 GLDM (gray level difference method) features, 11 GLRLM (gray level run length matrix) features and 59 LBP (local binary patterns) features. Finally, in classification phase, suspicious areas (detected in segmentation phase) have been classified in normal or abnormal (mass) tissue. For this purpose, seven different classifiers (support vector machine, naive bayes classifier, k-nearest neighbor, logistic regression, decision trees, random forest and multilayer perceptron neural network) have been tested by using 10-fold cross validation procedure.

II. DATABASE

In this study we used images taken from the mini-MIAS database. The database contains 322 digitised films. It also includes radiologist's "truth"-markings on the locations of any abnormalities that may be present (ie. data about positions and aproximate radii of masses are available). The database has been reduced to a 200 micron pixel edge and padded/clipped so that all the images are 1024x1024.

Mini-MIAS database provides information about character of the background tissue for each mammogram

(fatty, fatty-glandular or dense-glandular). This information we use in segmentation phase where we perform detection of suspicious regions by using different threshold functions for different type of background tissue. Database also provides information about class of abnormality present (microcalcification, well-defined masses, spiculated masses, ill-defined masses, architectural distortion, asymmetry). In radiology image the microcalcifications are seen as a very small (of size 0.1–1 mm) bright surfaces usually clustered in groups. Recognizing microcalcifications is not easy in many cases, but there are successful methods for their segmentation [4]-[5]. Unlike microcalcifications, other abnormality types are larger and not clustered, hence different algorithms are required for segmentation. In this study we will focus on detecting all abnormalities except microcalcifications. This reduces our database from 322 images to 298 images (after removing all mammogram having microcalcifications present). 92 masses are present on 89 images while other 209 images have no abnormality present.

III. IMAGE PREPROCESSING

A. Contrast Enhancement

In order to improve contrast, in this study we used Contrast-limited adaptive histogram equalization (CLAHE) algorithm. This algorithm operates on small regions in the image called tiles by enhancing their contrast. The neighbouring tiles are then combined using bilinear interpolation to eliminate artificially induced boundaries. The experimental results of enhancement on digital mammogram using CLAHE have been reported [6].

B. Filtering

An image is often corrupted by noise during its acquisition or transmission. In order to remove noise, within this study we used 2D adaptive median filtering. This algorithm performs spatial processing to determine pixels affected by impulse noise by comparing their value with neighbouring pixels. A pixel that is different from a majority of its neighbors is labeled as impulse noise and replaced with median value of neighbouring pixels. The size of the neighborhood (D) is adjustable (in this paper we determine neighborhood size thought optimization procedure – section V).

IV. IMAGE SEGMENTATION

A. Background and Pectoral Muscle Removal

The goal of the segmentation phase is to detect suspicious areas (potential masses) within mammograms. In order to detect suspicious areas, pectoral muscle and any artifact present outside the breast area should be removed first. Background is removed by finding the largest area of connected non-zero pixels (non-black pixels) and then setting all other pixel to zero (Fig. 1).

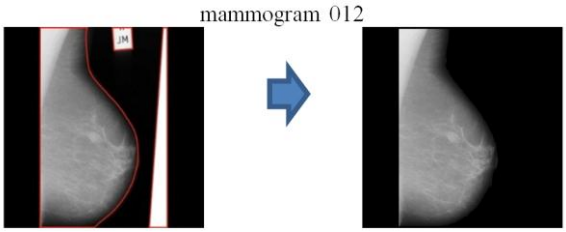


Figure 1. Background removal: selected areas of connected non-zero pixels (left), mammograms with removed background (right)

Pectoral muscle is removed by a procedure that uses the fact that pectoral muscle and a central part of the breast are usually denser (brighter) than the rest of the breast. Therefore, pectoral muscle and a central part of the breast can be extracted by applying local threshold operation with appropriate threshold value (c) which we optimize through the optimization procedure (section V).

Fig. 2(a) shows binary version of extracted pectoral muscle and a central part of the breast. Result of multiplication this binary image and original mammogram with removed background is shown in Fig. 2(b). In order to separate pectoral muscle from the central part of the breast shown in Fig. 2(b), it is first necessary to remove pectoral muscle from the image [7].

By subtracting isolated central tissue image (Fig. 2(c)) from the thresholded image (Fig. 2(b)) we get isolated pectoral muscle (Fig. 2(d)). Finally, mammogram with removed background and pectoral muscle is obtained by subtracting isolated pectoral muscle image (Fig. 2(d)) from the original mammogram with removed background. Mammogram depicted in Fig. 2(e) is used for the detection of suspicious regions as described in the following section.

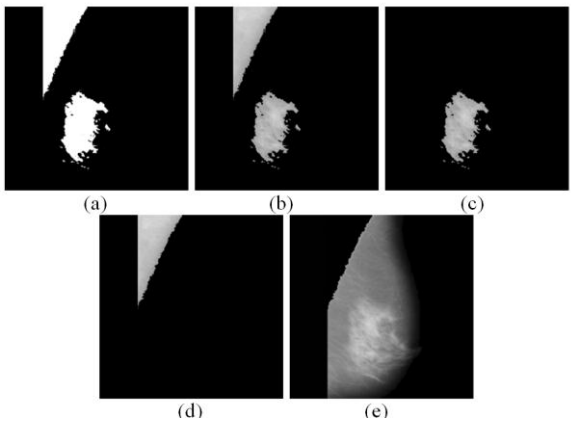


Figure 2. (a) Binary version of mammogram after local thresholding, (a) Original mammogram after local thresholding, (c) Extracted central tissue, (d) Isolated pectoral muscle, (e) Image with removed background and pectoral muscle

B. Detection of Suspicious Regions

Image shown in Fig. 2(e) is used as input for detection of suspicious regions. At this stage we detect suspicious areas separating them from the rest of the breast. Suspicious area (potential mass) is usually brighter than its surroundings, has almost uniform density, has a regular shape with varying size, and has fuzzy boundaries

[8]. At this stage we try to detect as much as possible masses even with a large number of false positives (which we eliminate by using classification algorithms at later stage).

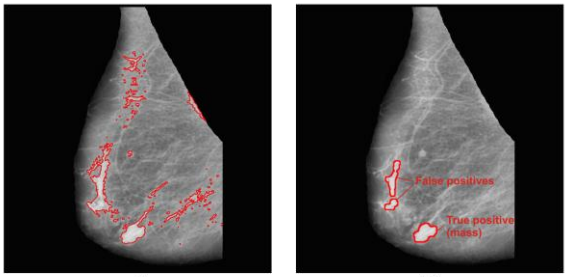
For separating suspicious areas we use global thresholding. We propose simple regression function as a threshold function:

$$Thres = a_1 I_{mean} + a_2 I_{\max} + a_3 I_{stdev} + a_4$$

$$I_{thres}(i,j) = \begin{cases} I(i,j) & I(i,j) \geq Thres \\ 0 & I(i,j) < Thres \end{cases} \quad (1))1)$$

where I_{mean} is average pixel value within breast area, I_{max} is maximum pixel value within breast area, I_{stdev} is standard deviation of pixel values within breast area, I_{thres} is image thresholded with *Thres* value, and a_1, a_2, a_3 and a_4 are parameters to be determined through optimization procedure (section V).

After thresholding with proposed threshold function (Fig. 3(a)), we perform morphological opening with disk-shaped structuring element in order to remove small objects (Fig. 3(b)).



a) b)

Figure 3. (a) Suspicious areas obtained after thresholding with threshold function, (b) Final suspicious areas obtained after morphological opening

By using threshold functions followed by morphological opening we were able to correctly segment 92% of masses with 4.14 false positives per image (FPpi).

V. PARAMETER OPTIMIZATION

In this section we describe optimization procedure that we used in order to maximize tumour detection sensitivity (number of correctly segmented masses) of preprocessing and segmentation phases.

For the three different types of background tissue (fatty, fatty-glandular and dense-glandular), we independently fit median filter neighborhood size (D), pectoral muscle isolation threshold value (c), threshold function parameters (a_1, a_2, a_3 and a_4) and size of the disk-shaped structuring element used for morphological opening (R). Optimization is performed by the use of simplex optimization method developed by John Nelder and Roger Mead [9]. Table I summarizes parameter values for fatty, fatty-glandular and dense-glandular background tissue obtained through optimization procedure.

TABLE I.
OPTIMIZED PARAMETER VALUES

Character of the background tissue	D	c	a_1	a_2	a_3	a_4	R
Fatty	3	155	0.22	0.56	0.46	1.54	10
Fatty-glandular	3	155	0.39	0.34	0.29	31.7	12
Dense-glandular	9	155	0.30	0.55	0.33	0.52	15

VI. FEATURE EXTRACTION

Features extraction is important step in breast cancer detection process. In this phase we describe every suspicious region by numerical attributes calculated by using different methodologies. In this paper we calculate 11 statistical features, 20 GLCM (gray level co-occurrence matrix) features, 5 GLDM (gray level difference method) features, 11 GLRLM (gray level run length matrix) features and 59 LBP (local binary patterns) features. Set of 106 attribute values have been calculated for every suspicious region and used as inputs for various classification algorithms which classify suspicious region into normal or mass tissue.

Basic statistical features are the simplest to calculate and represent the basic statistic of suspicious region. Table II summarizes statistical features calculated in this study.

TABLE II.
11 STATISTICAL FEATURES

Feature No.	Feature name
f1	Minimum
f2	Maximum
f3	Average
f4	Standard deviation
f5	Median
f6	Range
f7	Commonest value
f8	Trimmed mean (50%)
f9	Skewness
f10	Kurtosis
f11	Size

The gray-level co-occurrence matrix (GLCM) is a statistical method of examining texture that considers the spatial relationship of pixels. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix. In this study GLCM is calculated in four directions 0°, 45°, 90°, and 135° and distances d=1, and averaging is done to make direction invariant. In order to reduce the influence of random noise on texture features, the number of gray levels was reduced to 8 (from 256) prior to the accumulation of the matrix. Table III summarizes GLCM features calculated in this study. Features f1-f13 are features proposed by Haralick [10], Soh proposed features f14-f18 [11] and features f19 and f20 are proposed by Clausi [12].

TABLE III.
20 FEATURES EXTRACTED FROM GLCM

Feature No.	Feature name
f12	Autocorrelation
f13	Contrast
f14	Correlation
f15	Cluster Prominence
f16	Cluster Shade
f17	Dissimilarity
f18	Angular Second Moment (Energy)
f19	Entropy
f20	Inverse Difference Moment (Homogeneity)
f21	Maximum Probability
f22	Sum of Squares: Variance
f23	Sum Average
f24	Sum Variance
f25	Sum Entropy
f26	Difference Variance
f27	Difference Entropy
f28	Information Measure of Correlation 1
f29	Information Measure of Correlation 2
f30	Inverse Difference Normalized
f31	Inverse Difference Moment Normalized

The Gray Level Difference method (GLDM) [13] is based on the occurrence of two pixels having a given absolute difference in gray level and being separated by a specific displacement δ. Four possible forms of the vector δ exist: (0,d), (-d,d), (d,0), and (d,-d) where d is the intersample spacing distance. In this study we calculated five GLDM features (see table IV). Intersample spacing distance d=1 is used and averaging over the four angular directions was computed.

TABLE IV.
5 FEATURES EXTRACTED FROM GLDM

Feature No.	Feature name
f32	Contrast
f33	Angular second moment
f34	Entropy
f35	Mean
f36	Inverse difference moment

TABLE V.
11 STATISTICAL FEATURES11 FEATURES EXTRACTED FROM GLRLM

Feature No.	Feature name
f37	Short run emphasis
f38	Long run emphasis
f39	Gray-level non-uniformity
f40	Run length non-uniformity
f41	Run percentage
f42	Low gray-level run emphasis
f43	High gray-level run emphasis
f44	Short run low gray-level emphasis
f45	Short run high gray-level emphasis
f46	Long run low gray-level emphasis
f47	Long run high gray-level emphasis

The gray level run length matrix (GLRLM) [14] is another matrix commonly used for calculation of texture measures. Gray level run length matrix is a two-dimensional matrix where rows represent different values of gray levels and columns represent different number of runs in a certain direction θ. In this study, averaging over the four angular directions (0˚,45˚,90˚,135˚) was computed and the number of gray levels has been reduced to 8. Table V summarizes 11 GLRLM features calculated in this study.

Local Binary Pattern (LBP) is a simple yet very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number. LBP operator [15]-[16] forms labels for the image pixels by thresholding the R x R neighborhood of each pixel with the center value and considering the result as a binary number. The histogram of these 2^P different labels (where P is number of neighbors) can then be used as a feature vector. In order to reduce length of the feature vector the term uniform patterns have been introduced. A local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa. In this study we used 3 x 3 neighborhood (R=3, P=8), and used only uniform patterns (all the non-uniform patterns are labeled with a single label). In this way we calculated 59 LBP features (f48- f106).

VII. CLASSIFICATION MODELS AND EVALUATION

In order to classify suspicious areas into normal or mass tissues, calculated features (f1-f106) are given as inputs to 7 different classifiers. For classification process we used:

1) *support vector machine* [17]: a kernel method which transforms the input data space to optimize the fit to the optimal hyperplane. The algorithm uses a radial basis kernel function, parameters γ=1/(number of attributes), C=1 and the sequential minimal optimization training.
2) *naive bayes classifier*: a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. For numeric attributes (as in our case) naive bayes classifier uses Gaussian distributions [18].
3) *k-nearest neighbor*: instance-based lazy learning algorithm [19] which predicts the target class that is dominant among k most similar learning examples (nearest neighbors) in the problem space.
4) *logistic regression*: simple classification algorithm used to model dichotomous outcome variables. In this model the log odds of the outcome is modeled as a linear combination of the predictor variables [20].
5) *decision tree*: C4.5 decision trees algorithm [21] classifies instances by sorting them down the tree from the root to leaf node, which provides the classification of the instance. At each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the normalized information gain.
6) *random forest* [22]: an aggregate of N_{trees} stochastically built decision trees, where each tree represents a partial solution of the prediction problem. Predicted class is the most dominant class among partial predictions.

7) *neural network*: multilayer perceptron consisting from one hidden layer. A nonlinear sigmoid function is used as the activation function for each neuron, both the hidden layer and the output layer. Learning was performed using the backpropagation algorithm with momentum [23].

First, we calculated accuracy of these seven different classification algorithms by using all 106 features as inputs. Then, we performed MRMR (Minimum redundancy maximum relevance) feature selection algorithm [24] for selection of the 25 most relevant features. Models have been tested by using 10-fold cross validation procedure. In order to compare classification results of a different data mining algorithms we calculated accuracy, sensitivity, specificity and area under the ROC curve (A_z).

In this study, segmentation phase provided database containing 84 positive (mass) examples and 1233 negative (normal tissue) examples thus, this dataset is imbalanced. Big imbalance in data can cause some classifiers to perform poorly [25]. When the majority examples heavily out-number the minority examples some classifiers tend to ignore the minority class. Problem of class imbalance we solve by using SMOTE (Synthetic Minority Oversampling Technique) [26]. Instead of deleting or duplicating random examples in the dataset, this algorithm generates synthetic examples using the existing minority examples. In every iteration of the cross validation process, we balance data by performing SMOTE algorithm on training examples.

VIII. RESULTS

Table VI summarizes classification results (accuracy, sensitivity, specificity and area under ROC curve) for classifiers trained with all 106 input features. Classifiers have been tested by performing 10-fold cross validation procedure. Among seven classifiers, multilayer perceptron (MLP) neural network gave the best result (A_z =0.902).

TABLE VI.
CLASSIFICATION ACCURACY, SENSITIVITY, SPECIFICITY AND AREA UNDER ROC CURVE FOR DIFFERENT CLASSIFIERS TRAINED WITH ALL 106 FEATURES

Classifier	Accuracy	Sensitivity	Specificity	A_z
Naive Bayes	0.798	0.726	0.804	0.830
Logistic regression	0.839	0.833	0.839	0.892
SVM	0.848	0.833	0.849	0.841
KNN	0.832	0.655	0.845	0.814
C4.5	0.877	0.667	0.892	0.835
Random forest	0.911	0.655	0.929	0.901
MLP	0.878	0.726	0.889	0.902

In order to improve classification accuracy we extracted 25 most relevant features by using MRMR algorithm (see table VII). This algorithm tends to select features which are most relevant to the class and have the least correlation between themself.

TABLE VII.
LIST OF 25 FEATURES SELECTED BY USING MRMR ALGORITHM

Feature No.	Feature group	Feature name
f1	Statistical	Minimum
f11	Statistical	Size
f13	GLCM	Contrast
f14	GLCM	Correlation
f17	GLCM	Dissimilarity
f19	GLCM	Entropy
f20	GLCM	Inverse Difference Moment-Homogeneity
f26	GLCM	Difference Variance
f27	GLCM	Difference Entropy
f28	GLCM	Information Measure of Correlation 1
f29	GLCM	Information Measure of Correlation 2
f30	GLCM	Inverse Difference Normalized
f31	GLCM	Inverse Difference Moment Normalized
f33	GLDM	Angular second moment
f34	GLDM	Entropy
f36	GLDM	Inverse difference moment
f37	GLRLM	Short run emphasis
f38	GLRLM	Long run emphasis
f39	GLRLM	Gray-level non-uniformity
f41	GLRLM	Run percentage
f45	GLRLM	Short run high gray-level emphasis
f47	GLRLM	Long run high gray-level emphasis
f57	LBP	LBP 10
f85	LBP	LBP 38
f106	LBP	LBP 59 (non-uniform patterns)

Table VIII summarizes classification results, obtained by performing 10-fold cross validation, for classifiers trained with reduced set of features (25 features).

TABLE VIII.
CLASSIFICATION ACCURACY, SENSITIVITY, SPECIFICITY AND AREA UNDER ROC CURVE FOR DIFFERENT CLASSIFIERS TRAINED WITH SELECTED 25 FEATURES

Classifier	Accuracy	Sensitivity	Specificity	A_z
Naive Bayes	0.844	0.798	0.848	0.885
Logistic regression	0.898	0.774	0.907	0.894
SVM	0.845	0.798	0.848	0.823
KNN	0.873	0.667	0.887	0.842
C4.5	0.841	0.690	0.851	0.848
Random forest	0.914	0.679	0.930	0.898
MLP	0.874	0.774	0.881	0.902

By comparing results from tables VI and VIII we can conclude that most of classifiers showed improved results (regarding the A_z value). According to the results from table VIII, highest accuracy was achieved with random forest algorithm (0.914), but this is not relevant since we have imbalanced dataset (sensitivity for this model is 0.679 which is not good enough). For the best performing classifier we choose MLP neural network since it showed the highest area under ROC curve value. This classifier, trained with reduced set of features, retained highest A_z value (A_z=0.902). The structure for this neural network consisted of an input layer with a number of input neurons equal to the number of features of the used set (25), one hidden layer with a variable number of neurons, and an output layer. A nonlinear sigmoid function is used as the activation function for each neuron, both the hidden layer and the output layer. Learning was performed using the backpropagation algorithm with momentum [23]. In order to find the appropriate number

of hidden neurons (6), and the values of learning rate (0.3) and momentum (0.2), a trial-and-error process was applied. The stopping criterion was defined as a maximum number or learning epochs (5000). MLP classifier achieved 77.4% sensitivity and 88.1% specificity (0.49 FPpi). The ROC curve for this classifier is shown in Fig. 4.

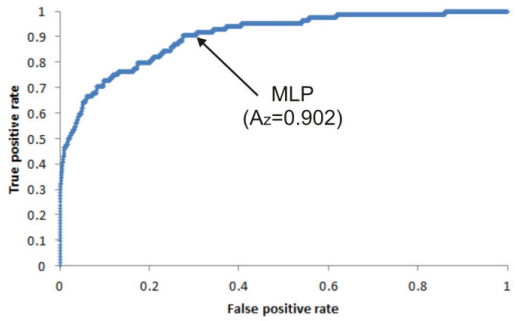


Figure 4. ROC curve for the MLP neural network classifier

IX. Discussion and Conclusion

We proposed CAD system for mass detection on digitized mammograms. By using proposed image enhacement and segmentation algorithms 92% of masses were correctly segmented with 4.14 false positives per image. In order to remove as much false positives as possible and keep most of detected masses, classification has been performed. For this purpose, various statistical, GLCM, GRLM, GLDM and LBP parameters have been extracted from suspicious areas. Among seven different classifiers tested by using 10-fold cross validation procedure, multilayer perceptron neural network showed the best result (highest area under ROC curve - A_z=0.902). This classifier achieved 77.4% sensitivity and 0.49 false positive per image.

Future work will include testing of the proposed CAD system on other datasets (we are currently collecting images from clinical center Kragujevac, Serbia) and focus on the discrimination between benign and malignant masses.

Acknowledgments

This study was funded by grants from Serbian Ministry of Science III41007 and ON174028.

References

[1] K. Bovis, S. Singh, J. Fieldsend, and C. Pinder, ”Identification of masses in digital mammograms with MLP and RBF nets,” in: Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks Com, pp. 342–347, 2000.

[2] H.D. Cheng, X.J. Shi, R. Min, L.M. Hu, X.P. Cai, and H.N. Du, ”Approaches for automated detection and classification of masses in mammograms,” *Pattern Recognition*, Vol. 39, pp. 646–668, 2006.

[3] V.D. Nguyen, D.T. Nguyen, T.D. Nguyen, and V.T. Pham, ”An Automated Method to Segment and Classify Masses in Mammograms,” International Journal of Electrical and Computer Engineering, Vol. 3, pp. 4-22, 2009.

[4] T. Stojic, I. Reljin, and B. Reljin, ”Adaptation of multifractal analysis to segmentation of microcalcifications in digital mammograms,” *Physica A*, Vol. 367, pp. 494–508, 2006.

[5] B. Reljin, Z. Milosevic, T. Stojiæ, and I. Reljin, ”Computer aided system for segmentation and visualization of microcalcifications in digital mammograms,” *Folia Histochemica et Cytobiologica*, Vol. 47, No. 3, pp. 525-532, 2009.

[6] I.K. Maitra, S. Nag, S.K. Bandyopadhyay, ”Technique for Preprocessing of Digital Mammogram,” Computer Methods and Programs in Biomedicine, Vol. 107(2), pp. 175-188, 2012.

[7] M. Radovic, M. Djokovic, A. Peulic, and Nenad Filipovic, ” Application of Data Mining Algorithms for Mammogram Classification,” 13 th IEEE International Conference on BioInformatics and BioEngineering, Chania, Greece, 10-13 November 2013.

[8] S.M. Lai, X. Li, and W.F. Biscof, ”On techniques for detecting circumscribed masses in mammograms,” *IEEE Trans. Med. Imaging*, Vol. l8, No. 4, pp. 377–386, 1989.

[9] J. Nelder, and R. Mead, ”A simplex method for function minimization,” *Computer Journal*, Vol. 7, No. 4, pp. 308-313, 1965.

[10] R.M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” *IEEE Transactions on systems, man and cybernetics*, vol. 3, no. 6, pp. 610-621, 1973.

[11] L.K. Soh, and C. Tsatsoulis, “Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices,” *IEEE Transactions on geoscience and remote sensing*, vol. 37, no. 2, pp. 780-795, 1999.

[12] D.A. Clausi, “An analysis of co-occurrence texture statistics as a function of grey level quantization,” *Canadian Journal of Remote Sensing*, vol. 28, no. 1, pp. 45–62, 2002.

[13] J.S. Weszka, C.R. Dyer, and A. Rosenfeld, ”A comparative study of texture measures for terrain classification,” *IEEE Trans. Syst., Man, Cybern.*, Vol. SMC-6, pp. 269–285, Apr. 1976.

[14] R.M.M. Galloway, “Texture analysis using gray level run lengths,” *Comput. Graphic. Image Processing*, Vol. 4, pp. 172–179, 1975.

[15] T. Ojala, M. Pietikäinen, and D. Harwood, ”A Comparative Study of Texture Measures with Classification Based on Feature Distributions,” *Pattern Recognition*, Vol 19, No.3, pp. 51-59, 1996.

[16] T. Ojala, M. Pietikäinen, and T. Mäenpää, ”Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 24, No.7, pp. 971-987, 2002.

[17] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 1995.

[18] H.J. George, and P. Langley, ”Estimating Continuous Distributions in Bayesian Classifiers,” In: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, pp. 338-345, 1995.

[19] D. Aha, and D. Kibler, ”Instance-based learning algorithms,” *Machine Learning*, Kluwer Academic Publishers, Boston, Vol. 6, pp. 37–66, 1991.

[20] S. le Cessie, and J.C. van Houwelingen, ”Ridge Estimators in Logistic Regression, ” *Applied Statistics*, Vol. 41, No. 1, pp.191-201, 1992.

[21] R. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[22] L. Breiman, ”Random forests,” *Machine Learning*, Vol. 45, No. 1, pp. 5–32, 2001.

[23] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Prentice Hall, New Jersey, USA, 1999.

[24] C. Ding, and H. Peng, “Minimum redundancy feature selection from microarray gene expression data,” *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 2, pp. 185-205, 2005.

[25] Y. Sun a, M.S. Kamel, A.K.C. Wong, Y. Wang, ”Cost-sensitive boosting for classication of imbalanced data,” *Pattern Recognition*, Vol. 40, pp. 3358–3378, 2007.

[26] N.V. Chawla, K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer, ”SMOTE: Synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, Vol. 16, No. 1, pp. 321–357, 2002.

Finite Element Model of Cochlea – Air Conduction and Bone Conduction

Velibor Isailović[*,**], Milica Nikolić [*,**], Žarko Milosević[*,**], Igor Saveljić[*,**], Dalibor Nikolić[*,**], Miloš Radović[*,**] and Nenad Filipović[*,**]

[*] Bioengineering Research and Development Center – BioIRC, Kragujevac, Serbia

[**] Faculty of engineering, University of Kragujevac, Serbia

velibor@kg.ac.rs, obradovicm@kg.ac.rs, zarko@kg.ac.rs, isaveljic@kg.ac.rs, markovac85@kg.ac.rs, mradovic@kg.ac.rs, fica@kg.ac.rs

***Abstract*— Cochlea is part of the inner ear. The role of the cochlea is to transform outer acoustic signal into electrical impulse which is further transmitted to the brain. Two important phenomena taking place inside the cochlea: transformation of outer acoustic signal into mechanical vibration and transforming of mechanical vibration into electrical impulse which is further transmitted to the brain. Cochlea has coiled shape like snail shell. Inside cochlea there are several parts, but from aspect of mechanics, the most significant parts are two fluid chambers: scala vestibuli and scala tympani [1] and elastic basilar membrane between them. The focus of this study is on mechanical part of the cochlea. Acoustic signal can get to the cochlea in two ways: through outer ear canal and trough the bones. First way is well known. The second one means that sound signal can get into cochlea traveling through the bones of skeleton. For example, acoustic excitation can be applied on some other part of the body and acoustic wave will come into cochlea through the bones. Objective of this work is to investigate effects of conduction of acoustic signals in these two ways. In this work simplified 3D finite element model of the cochlea was used. We assumed that shape of the cochlea doesn't affect on mechanical response of the basilar membrane [6].**

I. INTRODUCTION

Investigation of the physical processes inside cochlea became very interesting when Bekesy in 1960 found that different sound frequencies cause different shapes of oscillations of basilar membrane and stimulate different nerves which further send different signals to the brain [7]. In figure 1 is presented simplified view of human cochlea. There are several important parts: two cochlear chambers: scala vestibuli and scala tympani, basilar membrane, oval window and round window.

In investigations of pure mechanical response of cochlea some parts like scala media can be neglected. Also it is known that coiled shape of the cochlea doesn't affect on mechanical response. Including those assumptions we can consider cochlea as straight model with two fluid chambers: scala vestibuli and scala tympani. Besides that, our model contains parts like oval window, round window and helicotrema. Oval window is placed on the beginning of scala vestibuli, round window at the end of scala tympani and helicotrema couples scala vestibuli and scala tympani at the apex of cochlea [2], [4], [6].

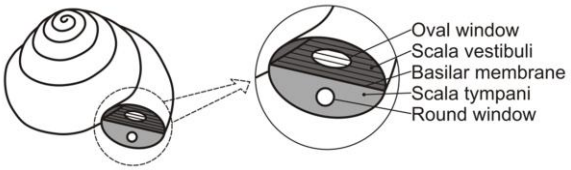


Figure 1. Simplified overview of human cochlea

For finite element analysis of the cochlea model in-house software was used. Theoretical base in that software is standard Newtonian dynamic equation for solid part and acoustic wave equation for fluid part.

In figure 3 are shown two models: uncoiled box model for air conduction and uncoiled tapered model for bone conduction.

II. METHODS

Model of the cochlea contain fluid and solid part. For solid part we use continuum mechanic approach based on principle of virtual work. Initially for fluid part of cochlea Navier – Stokes equation was used. But, we found that better approach for analysis of fluid domain of cochlea is to use acoustic wave equation to describe them [4].

Acoustic wave equation is defined as:

$$\frac{\partial^2 p}{\partial x_i^2} - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0 \qquad (1)$$

where is p pressure of fluid inside the chambers, x_i are spatial coordinates in Cartesian coordinate system, c is speed of sound, and t is time.

For numerical discretization of this equation we use standard Galerkin procedure [3],[5]. In that sense acoustic wave equation can be presented in matrix formulation:

$$Q\ddot{p} + Hp = 0 \qquad (2)$$

where Q is the acoustic inertia matrix, and H represents acoustic stiffness matrix.

Solid motion was described with Newtonian dynamics equation:

$$M\ddot{U} + B\dot{U} + KU = F^{ext} \qquad (3)$$

In equation (3) M, B and K are mass, damping and stiffness matrix, respectively.

Physical size and mass of the basilar membrane are constants in the model. It is not so in reality and in order to match place – frequency mapping, value of stiffness should be various along the basilar membrane. Value of stiffness as a function of distance from the base is (Ni 6):

$$E(x) = \frac{4\pi^2 f_B^2(x) A \rho \quad 1-\nu^2}{\beta^4 I} \quad (4)$$

where is: f_B - fundamental bending frequency, A - cross-sectional area of the basilar membrane, ρ - density, ν - Poisson's ratio, β - coefficient depends of boundary conditions and I - second moment of inertia.

In the modal analysis damping matrix could be included inside the stiffness matrix as a complex, imaginary part, so equation (3) could be written in the next form:

$$M\ddot{U} + K(1+i\eta)U = F^{ext} \quad (5)$$

where is η hysteretic damping ratio. This value can be expressed as a function of distance of material point from base [6]:

$$\eta = \frac{2\omega\xi_0 e^{\frac{x}{l}}}{\omega_B} \quad (6)$$

where is: ω - driving frequency, ξ_0 - damping ratio, x - distance from the base, l - natural frequency length scale and ω_B - natural frequency at the base.

For solving these equations the fluid – structure interaction with strong coupling (Filipovic et al, 2009) was used. Strong coupling means that solution of solid element in the contact with fluid has impact on the solution of fluid element. Coupling was achieved by equalization of normal pressure gradient of the fluid with normal acceleration of the solid element in contact, as it is expressed in equation (7).

$$n \cdot \nabla p = \rho n \cdot \ddot{u} \quad (7)$$

For mechanical model of the cochlea we defined system of coupled equation (8):

$$\begin{bmatrix} M & 0 \\ -\rho_f R & Q \end{bmatrix} \begin{Bmatrix} \ddot{U} \\ \ddot{p} \end{Bmatrix} + \begin{bmatrix} K(1+i\eta) & -S \\ 0 & H \end{bmatrix} \begin{Bmatrix} U \\ p \end{Bmatrix} = \begin{bmatrix} F \\ q \end{bmatrix} \quad (8)$$

where R and S are coupling matrices.

The solutions for displacement of the basilar membrane and pressure of the fluid in chambers were assumed in the following form:

$$U = A_U \sin(\omega t + \alpha)$$
$$p = A_p \sin(\omega t + \alpha) \quad (9)$$

In equation (9) A_U and A_p represent amplitudes of the displacement and the pressure, respectively. The circular frequency is ω, t is time and α is phase shift.

When displacement and pressure solution were substituted in the equation (8) we have system of the linear equations that can be solved (10):

$$\begin{bmatrix} K(1+i\eta) - \omega^2 M & -S \\ -\rho_f R & H - \omega^2 Q \end{bmatrix} \begin{Bmatrix} A_U \\ A_p \end{Bmatrix} = \begin{bmatrix} 0 \\ q \end{bmatrix} \quad (10)$$

This concept is very similar to the concept shown in literature [4], [6]. But, those authors have implemented this theory using Matlab, and they can solve just uncoiled models. Using our software it can be possible to model very complex geometry (for example, geometry obtained by images from medical devices, like computer tomography scanners).

III. RESULTS AND DISCUSSION

The first model which we have developed is box model with simplified geometry. In that model we have just basilar membrane and two fluid chambers: scala vestibuli and scala tympani (Fig. 2). This model was developed to model air conduction.

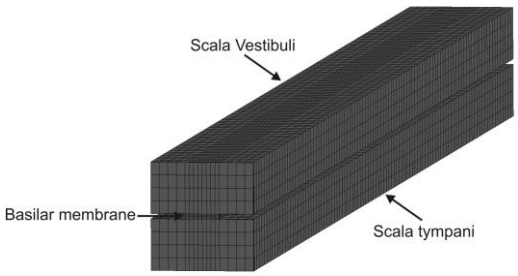


Figure 2: Air conduction finite element model

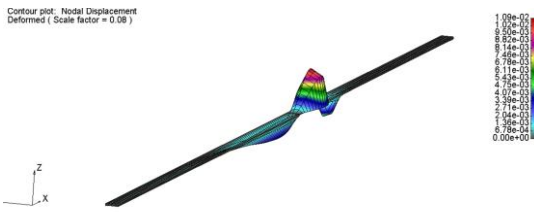


Figure 3: Air conduction response of the basilar membrane for excitation frequency of 1 kHz

Here we used prescribed velocity as excitation in the fluid domain. Excitation frequency was 1 kHz. In figure 3 is presented response of basilar membrane for that case. Presented results are normalized.

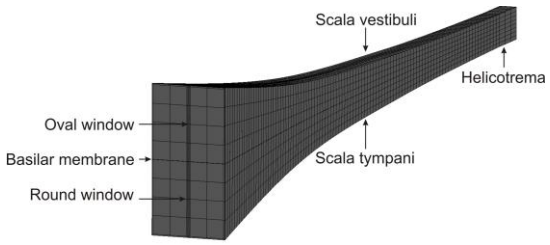


Figure 4: Bone conduction finite element model

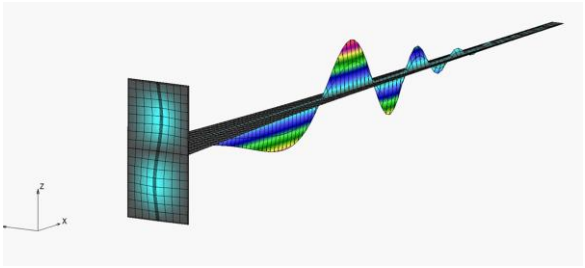


Figure 5: Bone conduction response of the basilar membrane for excitation frequency of 1 kHz

In second case we have a little bit complex model with basilar membrane, two fluid chambers, round and oval window and helicotrema (Fig. 4). In this model as excitation we used prescribed displacement on the edges of model to simulate excitation through the bones. The frequency of excitation was 1 kHz, as in previous model. Due to applied excitation, we have moving of whole system, but because of presence of elastic membranes on oval and round window, here have appeared oscillations of basilar membrane which are not in phase with prescribed excitation. The response of basilar membrane in case of applied prescribed displacement (bone conduction) is shown in figure 5.

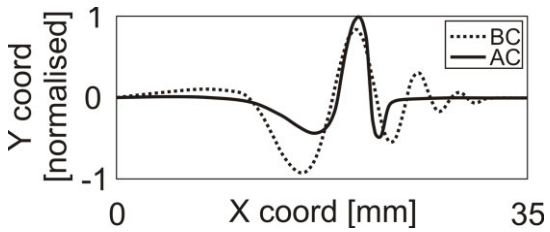


Figure 6: Air Conduction (AC) vs Bone Conduction (BC) response of the basilar membrane for excitation frequency of 1 kHz

In figure 6 are shown responses of basilar membrane for both cases, air conduction and bone conduction, with the same excitation frequency of 1 kHz. It shows that we will hear some tone in the same way regardless of the manner in which that reached the middle ear: by air conduction or by bone conduction. We have good agreement between these two results and they are also in good agreement with data from literature.

IV. CONCLUSION

Two simplified models of the cochlea were developed. We used those two models to investigate effects of air conduction and bone conduction of pure sound. Mechanical response of the cochlea was obtained by using in-house software PAK. Frequency domain analysis was applied in both cases.

Next step in modeling of the cochlea will be using real geometry of the cochlea obtained from CT scanners, because real geometry of the cochlea differs from patient to the patient. Also, those models will be analyzed with time domain solver to investigate response of the cochlea with arbitrary input excitation like white noise. Models of the cochlea can serve as a good support for clinicians in improving detection of the hearing disorders and losses.

ACKNOWLEDGMENTS

This work was supported in part by grants from Serbian Ministry of Education and Science III41007, ON174028 and FP7 ICT SIFEM 600933.

REFERENCES

[1] C.R. Steele (1987), "Cochlear Mechanics". In *Handbook of Bioengineering*, R. Skalak and S. Chien, Eds.,pp. 30.11–30.22, McGraw-Hill, New York.

[2] R. Nobili,F. Mommano, and J. Ashmore (1998), "How well do we understand the cochlea?". *TINS*, 21(4), pp.159–166.

[3] N. Filipovic, M. Kojic, R. Slavkovic, N. Grujovic, M. Zivkovic (2009), PAK, *Finite element software*, BioIRC Kragujevac, University of Kragujevac, 34000 Kragujevac, Serbia.

[4] S.J. Elliott, Ni G, Mace BR, Lineton B. (2013), "A wave finite element analysis of the passive cochlea". *The Journal of the Acoustical Society of America,* vol. 133, issue 3, p. 1535

[5] N. Filipovic, S. Mijailovic, A. Tsuda, & M. Kojic (2006), "An Implicit Algorithm within the Arbitrary Lagrangian-Eulerian Formulation for Solving Incompressible Fluid Flow with Large Boundary Motions". *Comp. Meth. Appl. Mech. Eng.*, Vol. 195, No. 44-47, pp. 6347-6361.

[6] Guangjian Ni (2012), "*Fluid coupling and waves in the cochlea*" – PhD thesis, University of Southampton, Faculty of engineering and the environment, Institute of sound and vibration research.

[7] von Békésy G. Experiments in hearing. New York: McGraw-Hill; 1960. 745 pp

Model-Based System for the creation and application of modified cloverleaf plate fixator

Nikola Vitković*, Mohammed Rashid*, Miodrag Manić*, Dragan Mišić*
Miroslav Trajanović*, Jelena Milovanović*, Stojanka Arsić**
* University of Niš, Faculty of Mechanical Engineering, Niš, Serbia
** University of Niš, Faculty of Medicine, Niš, Serbia
vitko@masfak.ni.ac.rs, miki_plast@yahoo.com, misicdr@gmail.com, miodrag.manic@masfak.ni.ac.rs,
miroslav.trajanovic@masfak.ni.ac.rs, jeka.milovanovic@gmail.com, stojanka@medfak.ni.ac.rs

***Abstract*— In healthcare systems there is a requirement to provide the best possible medical treatment for the patient, and that involves application of different procedures conducted by various experts in the field of medicine and other connected disciplines (engineers, software developers, managers, etc). For the medical treatment of the bone fractures various types of internal and external fixation techniques are used. The important components of these techniques are the fixators which enable fixation of broken bone parts. In this paper a new method for the customization of the modified cloverleaf plate fixator for the proximal part of the humerus bone is introduced. Furthermore, the application of Model-Based System Engineering (MBSE) for the modeling of the customization process is also presented. The main goal of this and future researches is to develop a complete system for the medical treatment of the humerus bone (and possible other human bones) which can improve patient general health and recovery process, and which will help doctors to improve their diagnostic and treatment skills in the field of orthopedic surgery.**

I. INTRODUCTION

In healthcare systems there is a requirement to provide the best possible medical treatment for the patient, and that involves application of different procedures conducted by various experts in the field of medicine and other connected disciplines (engineers, software developers, managers, etc). In the field of orthopedic surgery, especially in the field of bone trauma (e.g. fractures) this requirement can be achieved by the application of proper fixation technique for the implantation of customized fixator into the patient's body [1]. The important components of this technique are geometrically precise and anatomically accurate models of human bones. With such bone models, it is possible to build customized fixators [1] using rapid prototyping technologies or perform preoperative planning procedures [2,3]. Becides fixators and its design and implantation techniques, it is important to define the whole process of patient treatment from the diagnostic procedures to the full recovery. This process can be considered as System Engineering (SE) process [4] which involves all the management and technical skills in order to achieve patient/doctor requirements. Possible requirements can be: implantation of proper (customized) fixator, fast recovery process, bone function is returned to the state before trauma, etc. The management of the SE implies planning of technical details, risk management, control of the overall process, etc. The technical part of the SE implies specification of the technical data, hardware and software definition, etc [5]. SE applied in orthopedic surgery is a complex process and in order to conduct it in proper way experts from various fields must be included, like: orthopedists, surgeons, engineers, software developers, managers, etc. The application of SE in medicine or healthcare is a special academic field and it is often called Healthcare Systems Engineering (HSE) or Healthcare Engineering (HE), as described in [6,7]. It involves all processes which are needed for the physical and mental impairments of the patients. The modeling tools which are used for the graphical and textual representation of the HSE processes are based on the UML (Unified Modeling Language) which is maintained by the OMG (Object Management Group) [8]. The UML is a modeling language which focus is on modeling software systems and included processes, but there are various modifications which enable application of this language in other fields (e.g. SE). One of these modifications is SysML (Systems Modeling Language) which is a adapted UML language for System Engineers [9], and also maintained by OMG. SysML is a tool which can describe Model-Based System Engineering (MBSE) which is model oriented system engineering architecture [10]. The definition of MBSE from the official OMG site is "the formalized application of modeling to support system requirements, design, analysis, verification and validation activities beginning in the conceptual design phase and continuing throughout development and later life cycle phases."[8]. This means that this systems are model centric and not document centric as it was in previous or traditional approach [10]. It is expected that MBSE becomes an important factor of System modeling in the System Engineering practice in the future and to include various aspects of the systems, like social, economical, environment and other [11].

In this research MBSE is applied for modeling a part of the orthopedic treatment process for the patient with bone fracture. The part of the system which is modeled is fixator customization process. The method for the definition of fixator geometry and topology is newly developed by the authors involved in this research, and it enables creation of customized fixators for the human bones. This is very important process because it directly affects the success of the patient's treatment [1]. The example of this procedure is presented on the creation of the modified internal cloverleaf plate fixator for the

humerus bone [12, 13]. The tool for modeling this system is a SysML-Lite sub-language which is not a standard UML language, but it is successfully used for the modeling of simpler systems, or for making a prototype of system (or parts of the system) [5], as is the case in this research. The main goal of this and future researches is to develop a complete system which can improve patient general health and recovery process, and which will help doctors to improve their diagnostic and treatment skills in the field of orthopedic surgery.

II. THE CASE STUDY

The bone trauma that is analysis in this research is a fracture of proximal part of a humerus bone. There are various types of bone fractures and they are defined by the adequate classification [14, 15]. The fixator which is used for the treatment of the fractures is a modified cloverleaf plate fixator. The process of patient's treatment for the fixation of the mentioned fracture(s) is presented in Fig. 1. This process can be considered as a framework process, and because of that not all sub-processes are presented, like ones defined in [16].

The four main processes are:

Diagnostic procedure - In this procedure the bone fracture is defined through the verbal communication with the patient and with the medical imaging procedure.

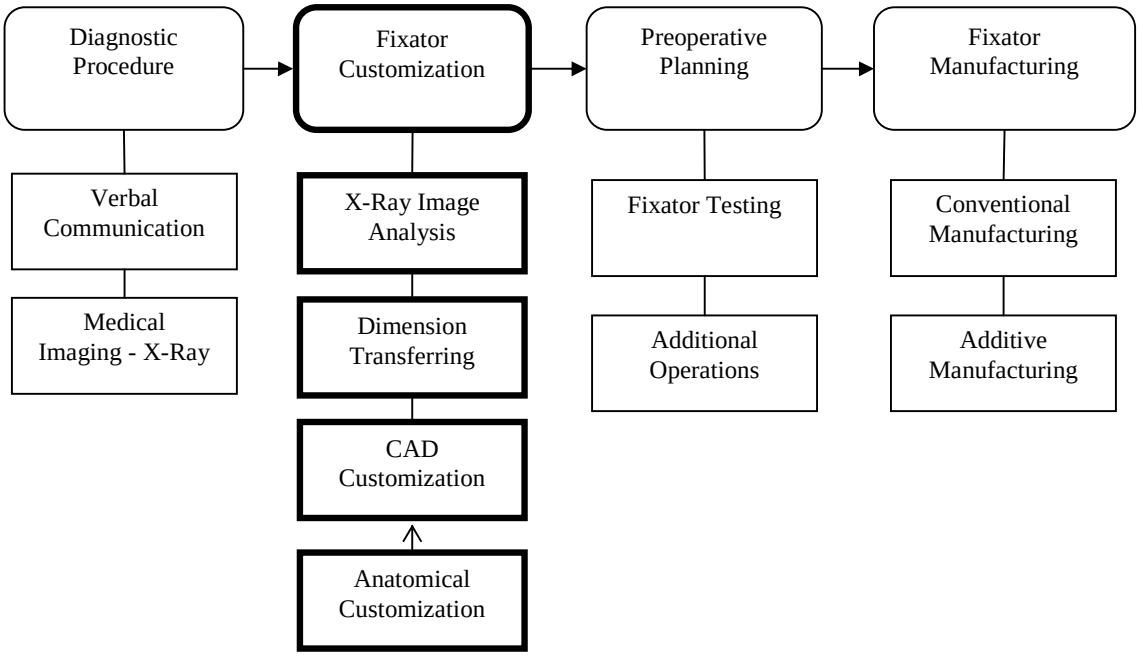


Figure 1. Scheme of the Customized fixator definition and manufacturing processes

In this case the procedure is conducted by the radiography (X-ray) scanning. To conduct a proper scanning it is desirable to scan the patient with some etalon placed near the patient. In this way proper scaling of the dimensions can be done. If etalon is not available, then X-ray scale for the scanning device must be known, which is always the case because it is supplied by the manufacturer of the radiograph.

Customization of the fixator - In this process geometrical model of the customized fixator for the specific patient is created. The input for this process is a digitized X-ray image of the patient's humerus bone. Based on the edge detection algorithm(s) applied in an open source software (e.g. GIMP in this case), it is possible to acquire the edge of the proximal part of the humerus bone and adequate values of dimensions measured in Anterior-Posterior plane of the humerus bone [2]. The dimensions which are measured are presented in the Fig. 2. There are two important dimensions RDmax (distal part of fixator) and RPmax (proximal part of fixator). These dimensions represent maximal distance from the detected edge to the Z axis of humerus body [12]. These dimensions enable creation of the profile curves with for the multisection feature in CATIA. The radius of this curves (part of the circle in this case) are limited by the values of RPmax and RD max respectively. There is another dimension which is important, and that is a rotational angle defined around Z axis of the fixator (angle of fixator rotation). This angle enables additional adjustment of the fixator position. In the dimensions transfer process adequate scaling of values is done, and the scaled values are stored in a textual file - Microsoft Excel. These values are the important input for the next process which is CAD Customization of the fixator's geometry. The acquired values are entered as values in CATIA and geometry of the parametric model of the fixator is adjusted to the patient's bone (dimensions and anatomy), as presented in the Fig. 2. This whole process of fixator customization is presented in the Fig. 3. This process can be improved by the application of the parametric bone model as described in [2]. The geometry of parametric bone model can also be adjusted to the measurements acquired for the specific patient. In that case it would be easier to adjust the fixator's geometry, because the bone geometry would be known. It should be noted that the geometry of fixator adapts to the geometry of the patient's bone based on only one 2D image, and not

on the basis of multiple images or 3D volumetric model (acquired by the CT or MRI), which means less radiation exposure. The outcome of this process is a 3D geometrical model of the customized fixator.

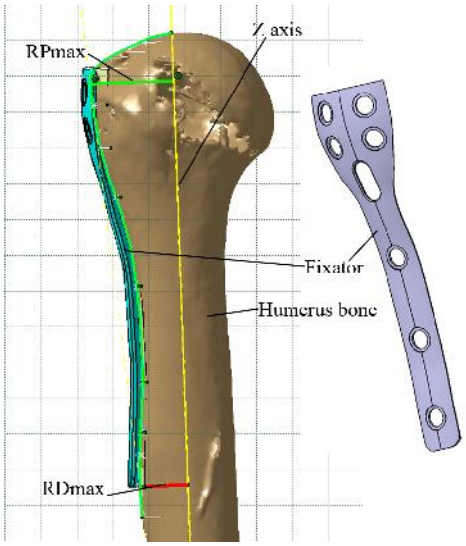


Figure 2. Scheme of the Customized fixator definition and production process

Preoperative planning - Based on the constructed fixator, medical practitioner (surgeon, orthopedist) plans the orthopedic intervention. This can be a very complex process and it involves planning all of the tasks which must be done in order to provide the best possible treatment for the patient. This process can be combined with previous process in order to better adjust the geometry and topology of the 3D model of the fixator.

Manufacturing of the fixator - The geometrical model of the customized fixator can be used in CAM for the conventional manufacturing processes (e.g. CNC machines), or for the manufacturing by the additive technologies, or by the combination of these two. In this case better solution for the manufacturing of the fixator will be additive technologies because of the complexity of the fixator's shape (free-form surface). The outcome of this process is a physical model of the customized fixator which can be implanted in the patient's body.

The main focus of this research is the process of fixator customization and the application of the MBSE for that process is presented in the next section of the paper.

III. MBSE WITH SYSML

The Systems Modeling Language (SysML) is a graphical modeling language for systems engineering applications. SysML is defined as a sub-language of the UML standard, and supports the specification, analysis, design, verification and validation of complex systems. SysML has been adopted by the Object Management Group (OMG) as OMG SysML (shorten SysML) and has evolved into standard for MBSE applications. "Model-Based Systems Engineering (MBSE) is a key enabling technology for Systems Engineers who seek to transition from traditional Systems Engineering processes that are document-based and code-centric to more effective processes that are requirements-driven and architecture-centric", as stated in [8]. These Systems Engineering activities include, but are not limited to, requirements analysis, functional analysis, performance analysis, and system architecture specification [8]. The SysML diagram contains five sub-diagrams and they are: package, requirement, behavior, parametric, and structure. Behavior diagram is represented by activity diagram, sequence diagram, state machine diagram and use case diagram. Structure diagram is defined by block definition diagram and internal block diagram. The detailed description of all of this diagrams is presented in [5, 8], but here only short definition is given.

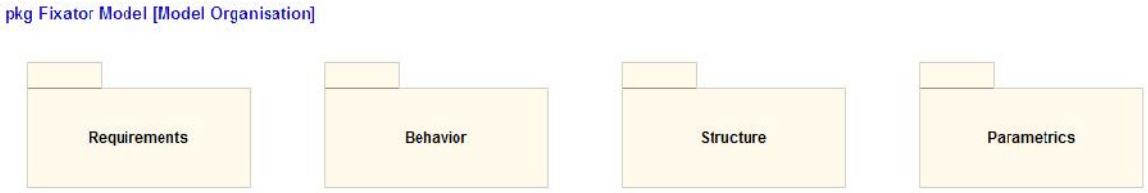
- Package diagram is the same as UML package diagram and it defines model elements organized in packages.
- Requirements diagram defines text based requirements and their relationships.
- Activity diagram is a modification of UML Activity diagram and it represents an activity flow in the order which activities are executed.
- Sequence diagram is the same as UML Sequence diagram and it defines the messages which are exchanged between systems or part of the systems.
- State machine diagram defines states of the entity(ies), which changes in a correspondence to the events - Same as UML State machine diagram.
- Use case diagram is standard UML diagram and it defines actions which are performed by the actors to the system in order to achieve defined goals.
- Block diagrams are the modification of UML class diagrams and they defined relationships between structural elements.
- Internal block diagram represents interconnection between parts of the block(s).
- Parametric diagram represents functional relationships between properties and its values.

In this research modified version of SysML called SysML-Lite is used, because it represents the prototype and not the complete and fully functional system which is tested in practice. The application of the system prototype in real clinical practice would give proper verification of the proposed system. The SysML-Lite diagram contains only six (6) diagrams and they are: package, requirement, activity, block definition, internal block, and parametric diagram. The tool which is used for modeling with SysML is Modelio SysML Architect Module. Modelio is Open Source software based on Eclipse environment [17]. The important diagrams which were created are presented in Fig. 4a - 4d, and they are:

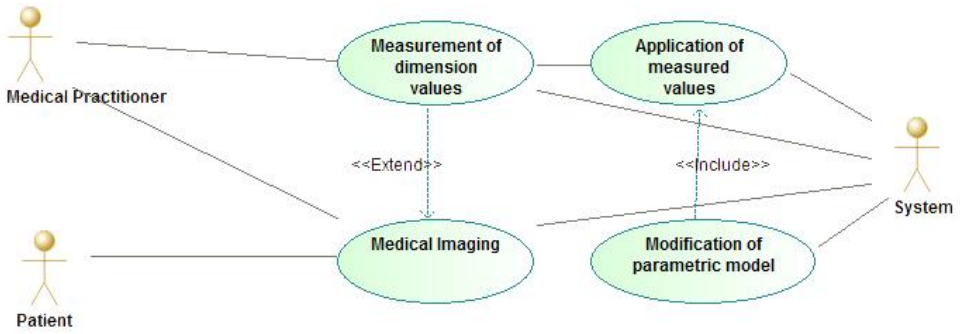
- Package diagram with defined packages for Requirements, Behavior, Structure and Parametrics.
- Use Case diagram for the customization of geometrical model of the fixator.
- Block Diagram for the Fixator Customization Context - All other blocks connected to the context are displayed.
- Internal block diagram for the fixator - In this diagram component of the fixator are presented together with attribute values.
- Parametric diagram for the internal block for the fixator with defined attributes.

- Requirement diagram for this process contains only one requirement and that is a proper adaptation of the fixator, so there is no need for the construction of that diagram for the presented case.

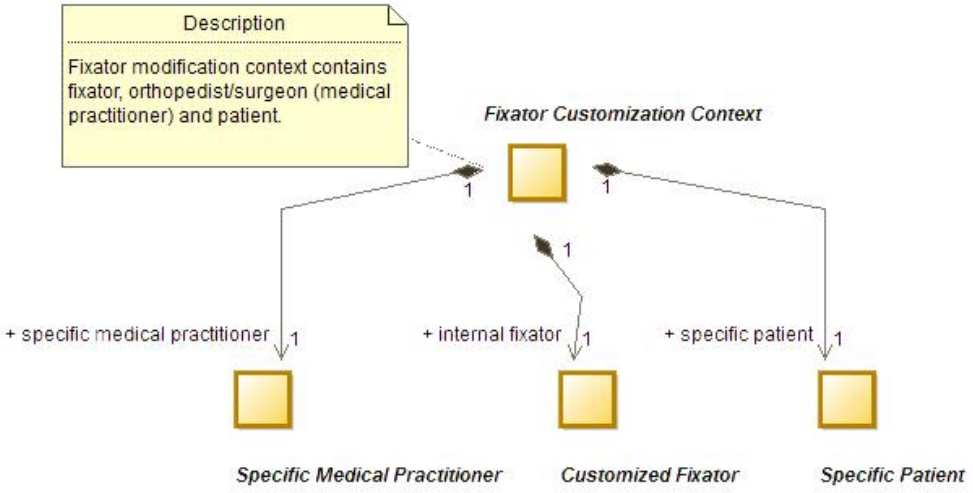
In the future work more diagrams will be created, some diagrams will be modified, and the whole business process (model) will be included.



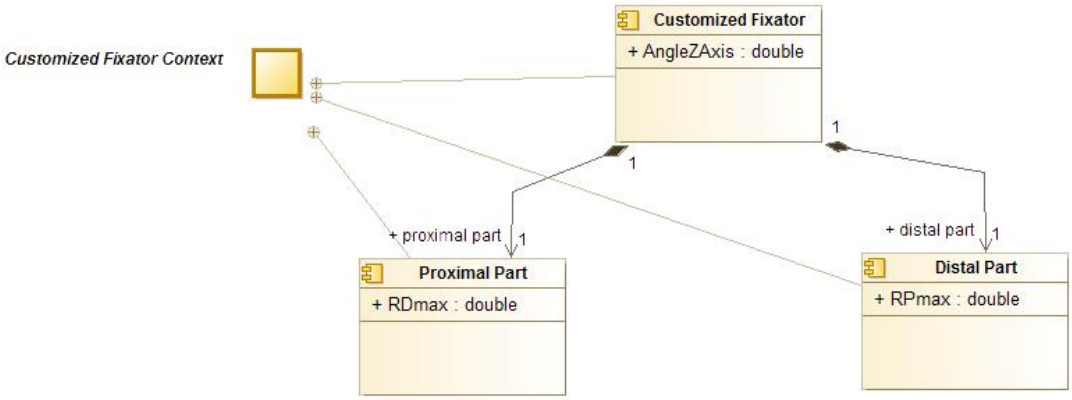
a) Package diagram - SysML-Lite



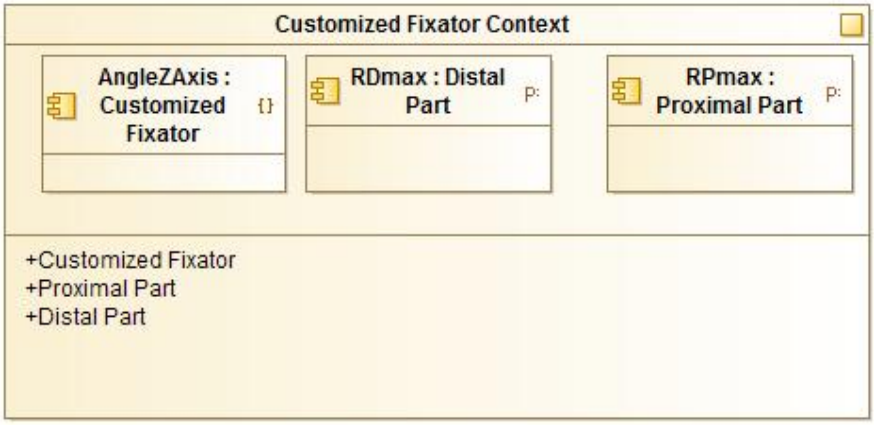
b) Use Case Diagram for Customization of fixator's geometrical model



c) Block Diagram for Fixator Customization Context



d) Internal block diagram for Customized Fixator Context



e) Parametric diagram for Customized Fixator Context block

Figure 3. SysML diagrams for the Customization of the modified cloverleaf plate fixator process

IV. CONCLUSION

In this research application of the MBSE for modeling the customization process of internal cloverleaf fixator for the humerus bone of the specific patient is presented. The process is based on newly developed method for the creation of the geometrical model of the customized fixator. This method is developed in order to enable creation of fixator geometrical model by the use of only one 2D X-ray image of the patient's bone, which means that patient will be exposed to the lesser radiation dose compared to volumetric scanning (e.g. CT).

The presented system is a prototype system, and in the future additional research will be conducted in order to create the complete system for the medical treatment of the patient with the bone trauma.

ACKNOWLEDGMENT

The paper presents the case that resulted from application of multidisciplinary research in the domain of bioengineering in real medical practice. The research project (Virtual Human Osteoarticular System and its Application in Preclinical and Clinical Practice) is sponsored by the Ministry of Science and Technology of the Republic of Serbia - project id III 41017 for the period of 2011-2014.

REFERENCES

[1] D. Stevanović, N. Vitković, M. Veselinović, M. Trajanović, M. Manić, M. Mitković, Parametrization of internal fixator by Mitkovic, International Working Conference ''Total Quality Management – Advanced and Intelligent Approaches'', 4th – 7th June, 2013., Belgrade, Serbia, 2013, pp. 541-544

[2] N. Vitković, J. Milovanović, N. Korunović, M. Trajanović, M. Stojković, D. Mišić, S. Arsić, Software System for Creation of Human Femur Customized Polygonal Models, Computer Science and Information Systems, Vol. 10, No. 3, 2013, pp. 1473-1497.

[3] E. Ramirez, E. Coto, Digital preoperative planning for long-bone fractures, Proceedings del X Congreso Internacional de Metodos Numericos en Ingenieria y Ciencias Aplicadas (CIMENICS 2010). Marzo, 2010, pp. TC 1-6.

[4] M. Linhares, A. Silva, R. Oliveira, Empirical Evaluation of SysML through Modeling of an Industrial Automation Unit, ETFA'2006 - 11th IEEE International Conference on Emerging Technologies and Factory Automation Prague, Czech Republic, 20-22 September 2006, pp 145-152.

[5] S. Friedenthal, A. Moore, R. Steiner, A Practical Guide to SysML The Systems Modeling Language, Morgan Kaufmann(Elsevier imprint), 225 Wyman Street, Waltham, MA 02451, USA, ISBN: 978-0-12-385206-9

[6] D. Sittig, H. Singh, A New Socio-technical Model for Studying Health Information Technology in Complex Adaptive Healthcare Systems, Oct 2010, doi: 10.1136/qshc.2010.042085

[7] A. Kushniruk, Evaluation in the design of health information systems: application of approaches emerging from usability engineering, Computers in Biology and Medicine, Vol. 32, No 3, 2002, pp. 141–149.

[8] Object Modeling Group web site, http://www.omg.org

[9] Ana. Ramos, J. Ferreira, J. Barcel, Model-Based Systems Engineering: An Emerging Approach for Modern Systems, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions, Vol. 42, No. 1, 2011, pp. 101-111

[10] J. Estefan, Survey of Candidate Model-Based Systems Engineering (MBSE) Methodologies, rev. B. Seattle, WA, USA: International Council on Systems Engineering (INCOSE). INCOSE-TD-2007-003-02. 2008

[11] C. Haskins, K. Forsberg, and M. Krueger, in Systems Engineering Handbook—A Guide for System Life Cycle Processes and Activities, Eds. San Diego, CA: Int. Council Syst. Eng., 2007.

[12] Proximal humerus 11-A3 Open reduction web site https://www2.aofoundation.org/

[13] D. Ruch, R. Glisson, A. Marr, G. Russell, J. Nunley, Fixation of three-part proximal humeral fractures: a biomechanical evaluation, Journal of Orthopaedic Trauma, Vol. 14. No. 1, 2000, pp. 36-40.

[14] Web site, http://www.shoulderdoc.co.uk/article.asp?article=1457#1_AO

[15] M. Sidor, J. Zuckerman, T. Lyon, K. Koval, F. Cuomo, N. Schoenberg, The Neer classification system for proximal humeral fractures. An assessment of interobserver reliability and intraobserver reproducibility, Journal of Bone and Joint Surgery Vol. 75, No. 12, 1993, pp. 1745-50.

[16] D. Mišić, M. Stojković, N. Vitković, M. Trajanović, M. Manić, N. Korunović, J. Milovanović, The concept of the information system for managing business processes of designing and manufacturing of osteofixation material, ICIST 2014 - Vol. 1 Regular papers, PC centar Magus, Zrenjanin, ISBN: 978-86-85525-14-8, 2014, pp. 10-16.

[17] Modelio web site, http://www.modelio.org

Decision Support System for Selection of the Most Suitable Biomedical Material

Dušan Petković*, Miloš Madić*, Goran Radenković*, Miodrag Manić*, Miroslav Trajanović*
*University of Niš, Faculty of Mechanical Engineering/Department for Production, IT and Management, Niš, Serbia
dulep@masfak.ni.ac.rs
madic@masfak.ni.ac.rs
rgoran@masfak.ni.ac.rs
miodrag.manic@masfak.ni.ac.rs
miroslav.trajanovic@masfak.ni.ac.rs

***Abstract*—Selection of the most suitable material for a given biomedical application is very complex, important and responsible task. A decision support system based on the use of method of multi-criteria decision making (MCDM) methods, named MCDM Solver, was developed in order to facilitate the selection process of biomedical materials and increase selection confidence and objectivity. In this paper, a MCDM problem which referes to the selection of the most suitable material for the compensation of the missing parts of the long bones was solved by using the developed MCDM Solver.**

I. INTRODUCTION

Biomedical materials are commonly characterized as materials used to build artificial organs, rehabilitation devices, or implants to replace natural body tissues [1]. Developing new and improved biomedical implants is seen as a complex design problem-solving activity and, in conjunction with demanding manufacturing constraints, utilizing the most appropriate materials (and materials combinations) presents many unique challenges [2]. Selecting an appropriate material for a given biomedical application is important from more points of view – medical, technological, and economic. Today, there is a large number of biomedical materials and manufacturing processes, each having its own properties, applications, advantages and limitations. Therefore, many difficult decisions need to be made while selecting a material for a specific biomedical implant. Decision makers have to consider a number of issues related to materials' mechanical, biological, chemical, physical technological and economic properties which to the greatest extent affect the quality and application of a biomedical product in a particular domain. Existence of correlations and contradictions among these properties makes the selection procedure more challenging and time consuming for decision makers. In order to select the most suitable biomedical material, the decision maker should have a complete understanding of the functional requirements of the product and a detailed knowledge of the considered criteria for a specific biomedical application. The unsuitable choice of a biomedical material may lead to a premature failure of the product, a need for repeated surgery, a cell death, chronic inflammation or other impairment of tissue functions as well as an extension of healing period and overall increasing of the costs [3]. Therefore, the designers in collaboration with medical specialists must identify and select the most suitable material for an implant device with the minimum possible cost and specific performance considerations.

The objectives and criteria in the material selection process are often in conflicts which involves certain trade-offs amongst decisive factors, such as desired properties, operating environment, production process, cost, market value, availability of supplying sources and product performance [4]. Only with a systematic and structured mathematical approach the best alternative for a specific engineering product can be selected.

The material selection problems with multiple non-commensurable and conflicting criteria can be efficiently solved using multi-criteria decision making (MCDM) methods. The MCDM methods have the capabilities to generate decision rules while considering relative significance of considered criteria upon which the complete ranking of alternatives is determined [5].

Various approaches have already been proposed by the past researchers to solve the material selection problem. Within the common used it can be listed Ashby approach [6], TOPSIS [7], ELECTREE [8], VIKOR [9], COPRAS [10], ANP [11], UTA method [12].

This paper presents the application of the developed software prototype i.e. decision support system (DSS) for the selection the most suitable biomedical material for bone implants which compensate the missing part of a long bone. Within the DSS, named MCDM Solver, a list with potential materials and their properties is created. Three MCDM methods are available for ranking the list of alternative materials, i.e. TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution), VIKOR (Više kriterijumska optimizacija – kompromisno rešenje) and WASPAS (Weighted Aggregated Sum Product Assessment). Based on the methods, the most appropriate material was selected and comparison of the ranking results was carried out.

II. APPLICATION OF THE MCDM METHODS FOR MATERIALS SELECTION

Generally, every MCDM problem starts with the decision/evaluation matrix [13],

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \cdots & \cdots & \cdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}$$

Where, m is the number of candidate alternatives (potential materials), n is the number of evaluation criteria (material properties) and x_{ij} is the performance of i th alternative with respect to j th criterion. Depending on the desired material property, a material performance can be assessed as quantitative number value or qualitative description based on the knowledge and experience of the decision maker, material designers and material users. Determination of the decision matrix is the first step of the material selection process. For example, if the first and second criteria are corrosion resistance and tensile strength, respectively, initial evaluation matrix for three alternative materials is as follow:

$$X = \begin{bmatrix} 7 & 630 \\ 10 & 550 \\ 9 & 655 \end{bmatrix}$$

Where, 7-10-9 and 630-550-655 are relative evaluation of the corrosion resistance and tensile strength of analyzed three materials, respectively. The next one is normalization of the performances in order to obtain dimensionless numbers ranged from 0 to 1. Normalization is followed by mathematical computing which can provide the base for ranking of the alternatives.

A. TOPSIS method

The TOPSIS method is proposed by Chen and Hwang [1]. The basic principle is that the chosen alternative should have the shortest distance from the positive ideal solution (PIS) and the farthest distance from the negative ideal solution (NIS). Therefore, this method is suitable for risk avoidance designer(s), because the designer(s) might like to have a decision which not only makes as much profit as possible but also avoids as much risk as possible [9]. The TOPSIS method has been used predominantly in materials selection due to its superior characteristics [15]. Pseudo algorithm of the TOPSIS method is shown in Figure 1.

B. VIKOR method

The VIKOR method was introduced as one applicable technique to implement within MCDM [16]. This method was developed for multi-criteria optimization of complex systems, which enjoys a wide acceptance [9]. It focuses on ranking and selecting from the alternatives with conflicting and different units criteria. In the VIKOR method, the compromise ranking is performed by comparing the measure of closeness to the ideal alternative, and compromise means an agreement established by mutual concessions. Pseudo algorithm of the VIKOR method is shown in Figure 2.

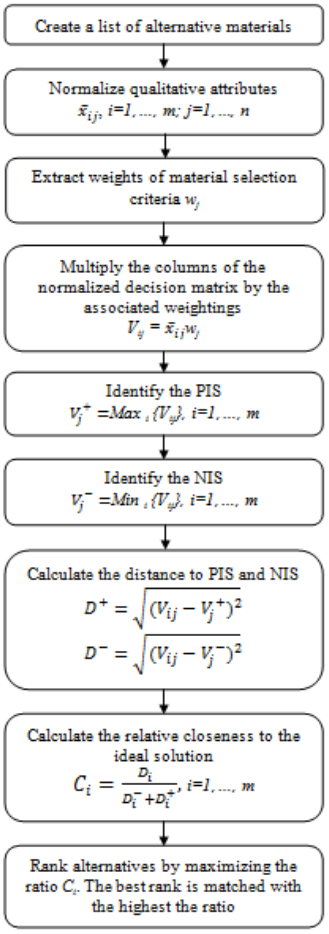


Figure 1. TOPSIS pseudo algorithm

C. WASPAS method

In order to increase ranking accuracy and reliability, a new methodology for optimization of weighted aggregated function was proposed. This method was named as the Weighted Aggregated Sum Product Assessment (WASPAS) and introduced firstly by Zavadskas et al. [17].

From the mathematical point of view, the WASPAS method presents a linear combination of weighted sum method (WSM) and weighted product method (WPM).

The application of the method at first requires linear normalization of the decision matrix, which is followed by weighting of the criteria. A general pseudo algorithm of the WASPAS method is shown in Figure 3.

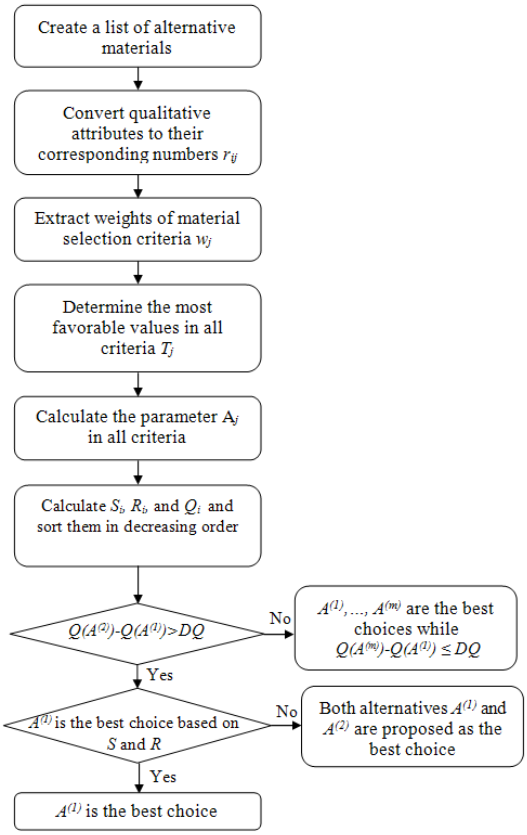


Figure 2. VIKOR pseudo algorithm

III. MDCM SOLVER PROTOTYPE FOR SOLVING BIOMEDICAL MATERIAL SELECTION

Decision making which involves a large number of variables (criteria, alternatives) requires advanced and comprehensive knowledge in the applied field. A lot of time is consumed in every selection process due to tedious calculations involved in evaluating each alternative with respect to the selection criteria. To eliminate these tedious calculations and ease out the material selection decision making process, a software prototype named MCDM Solver is developed. This desktop application is developed in Microsoft Visual Studio 2008 environment, using C# as a programming language.

The developed MCDM Solver integrates the users' requirements with the technical requirements and can be used to select the most appropriate biomedical material for a given application based on the selected requirements, as considered in the present research work.

Namely, this study is aimed to select the most suitable biomedical material for bone implants which compensate the missing part of a long bone. The main role of the implant is to replace the human bone in term of the function and aesthetic. Therefore, the implant material must have the properties so close to the properties of the missing bone, i.e. excellent biocompatibility, mechanical strength, wear and corrosion resistance. The compressive strength of compact bone is about 140 MPa, and the elastic modulus is about 14 GPa in the longitudinal direction and about 1/3 of that in the radial direction. These values are modest compared to most engineering materials. However, live healthy bone is self-healing and has a great resistance to fatigue loading. The implant is fixed to the surrounding bone structure by screws.

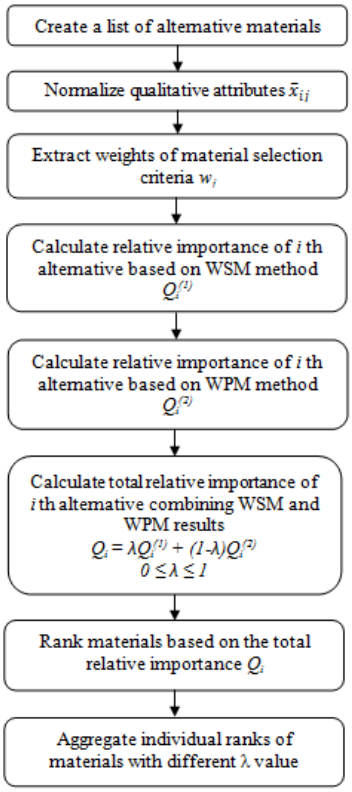


Figure 3. WASPAS pseudo algorithm

In order to select the most suitable bone implant biomaterial, requirements (criteria for selection) such as tissue tolerance, corrosion resistance, mechanical behavior, elastic compatibility, weight and cost have been considered. The initial list of potential biomedical materials with criteria and their performances is defined based on the previous similar studies [15, 18].

Figure 4 shows the screen shot of the developed MCDM Solver where candidate materials, criteria, weights, desired properties of the materials (target values) and available MCDM methods are shown.

IV. RESULTS AND DISCUSSION

Results of the proposed rankings of the biomedical materials for bone implant are presented in Table I. As could be seen from the table, application of different methods gives different ranking order. The best ranked material according to both TOPSIS and VIKOR methods is material number 6 (Co-Cr wrought alloy), while material number 8 (Ti-6Al-4V) is proposed as the best solution by using the WASPAS method. The WASPAS method proposed material number 6 as the second ranked material. Therefore, by applying an aggregation technique Co–Cr wrought alloy is proposed as the most appropriate biomedical material for the bone implant.

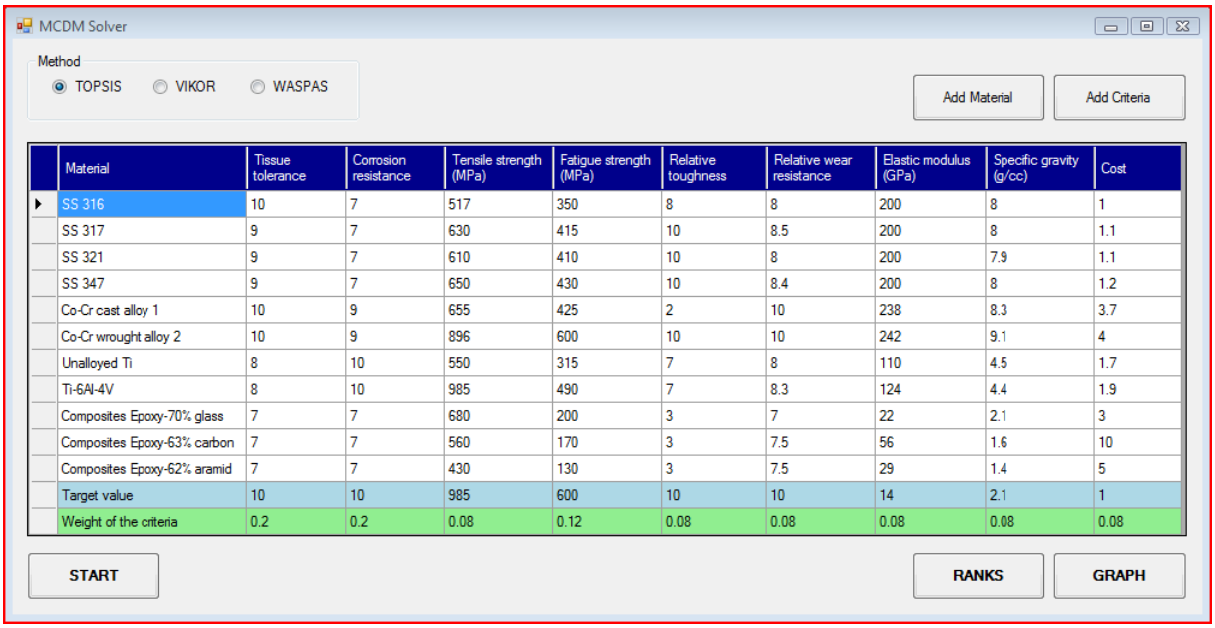
TABLE I. RANKING RESULTS OF THE BONE IMPLANT MATERIAL BASED ON THE TOPSIS, VIKOR AND WASPAS METHODS

No.	Material	TOPSIS	VIKOR	WASPAS
1	SS 316	5	5	7
2	SS 317	7	7	3
3	SS 321	8	8	6
4	SS 347	6	6	5
5	Co-Cr cast alloy 1	3	2	9
6	Co-Cr wrought alloy 2	1	1	2
7	Unalloyed Ti	4	4	4
8	Ti-6Al-4V	2	3	1
9	Composites Epoxy-70% glass	9	9	8
10	Composites Epoxy-63% carbon	11	11	11
11	Composites Epoxy-62% aramid	10	10	10

On the other hand, all three methods consistently yielded unalloyed Ti, composites epoxy-62% aramid and composites epoxy-63% carbon as the 4, 10 and 11 ranked materials, respectively.

It is also noticed that there is no total match between the methods but a very strong correlation between the TOPSIS and VIKOR methods (Spearman's rank correlation coefficient - 99%) is evident. Spearman's rank correlation coefficient also shows 71% between the TOPSIS and WASPAS methods and 64% between the VIKOR and WASPAS methods.

Taking into account pretty different ranking results obtained by the WASPAS method, an additional analysis is carried out. Hence the effect of varying values of the parameter λ on the rankings of the considered material selection is graphically presented in Figure 5. It is clearly visible that the rankings of the best material alternative (material number 8) remains unaffected for different values of the parameter λ.



Material	Tissue tolerance	Corrosion resistance	Tensile strength (MPa)	Fatigue strength (MPa)	Relative toughness	Relative wear resistance	Elastic modulus (GPa)	Specific gravity (g/cc)	Cost
SS 316	10	7	517	350	8	8	200	8	1
SS 317	9	7	630	415	10	8.5	200	8	1.1
SS 321	9	7	610	410	10	8	200	7.9	1.1
SS 347	9	7	650	430	10	8.4	200	8	1.2
Co-Cr cast alloy 1	10	9	655	425	2	10	238	8.3	3.7
Co-Cr wrought alloy 2	10	9	896	600	10	10	242	9.1	4
Unalloyed Ti	8	10	550	315	7	8	110	4.5	1.7
Ti-6Al-4V	8	10	985	490	7	8.3	124	4.4	1.9
Composites Epoxy-70% glass	7	7	680	200	3	7	22	2.1	3
Composites Epoxy-63% carbon	7	7	560	170	3	7.5	56	1.6	10
Composites Epoxy-62% aramid	7	7	430	130	3	7.5	29	1.4	5
Target value	10	10	985	600	10	10	14	2.1	1
Weight of the criteria	0.2	0.2	0.08	0.12	0.08	0.08	0.08	0.08	0.08

Figure 4. Screen shot of the MCDM Solver - initial step

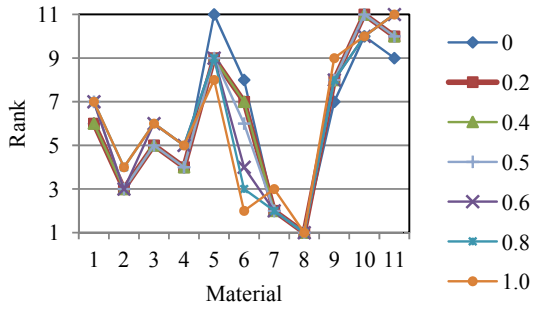


Figure 5. Variations of the materials rank depending on the values of parameter λ

V. CONCLUSION

The selection of a the most suitable material for a given biomedical application is a complex task due to limited knowledge about specific material properties, as well as complexity of the requirements which involve a large number of additional considerations. The use of the MCDM methods is proved to be a powerful tool for making genuine and objective decisions while selecting the most suitable biomedical material.

In this paper, the application of the TOPSIS, VIKOR and WASPAS methods for the selection of the most suitable bone implant material is demonstrated. A software prototype, named MCDM Solver, is developed to automate this material selection decision making process. It is important to highlight that the developed

software can be wider applied within the material selection field with the possibility to add new alternatives (materials) and criteria for the evaluation of alternatives.

Although there is no full match of results, the optimum choice of materials is clear. Additionally, it can be noticed very strong correlation between the TOPSIS and the VIKOR method.

ACKNOWLEDGMENT

This paper is a result of the projects III41017 and TR35034 supported by the Ministry of Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] D. Petković, G. Radenković, M. Mitković, "Fractographic investigation of failure in stainless steel orthopedic plates", *Facta Universitatis: Series Mechanical engineering*, vol. 10, no 1, pp. 7–14, 2012.

[2] A. Jahan, K. L. Edwards, *Multi-criteria Decision Analysis for Supporting the Selection of Engineering Materials in Product Design*, Butterworth-Heinemann, 2013.

[3] D. Petković, F. Živić, G. Radenković, M. Trajanović, and M. Manić, "Coating: a way to improve biomedical properties of AISI 316L stainless steel", *35 ICPE*, Kopaonik, pp. 167-174., 2013.

[4] P. Chatterjee, S. Chakraborty, "Material selection using preferential ranking methods", *Materials and Design*, vol. 35, pp. 384-393, 2012.

[5] D. Petković, M. Madić, G. Radenković, "Gear material selection using WASPAS method", *3rd International Congress Science and Management of Automotive and Transportation Engineering - SMAT 2014*, Craiova, Romania, pp. 45-48., 2014.

[6] M. F. Ashby, Y.J.M. Brechet, D. Cebon, and L. Salvoc, "Selection strategies for materials and processes", *Materials & Design*, vol. 25, pp. 51-67., 2004.

[7] D-H. Jee, and K-J.Kang, "A method for optimal material selection aided with decision making theory", *Materials & Design*, vol. 21, pp. 199-206., 2000.

[8] A. Shanian, O. Savadogo, "A material selection model based on the concept of multiple attribute decision making", *Materials & Design*, vol. 21, pp. 199-206., 2000.

[9] S. Opricovic, G. H. Tzeng, "Compromise solution by MCDM methods: a comparative analysis of VIKOR and TOPSIS*", Eur J Oper Res*, vol. 156, pp. 445–455., 2004.

[10] P. Chatterjee, V. M. Athawale, and S. Chakraborty, "Materials selection using complex proportional assessment and evaluation of mixed data methods", *Materials & Design*, vol. 32, pp. 851-860., 2011.

[11] A.S. Milani, A. Shanian, C. Lynam, T. Scarinci, "An application of the analytic network process in multiple criteria material selection", *Materials & Design*, vol. 44, pp. 622-632., 2013.

[12] V. M. Athawale, R. Kumar, S. Chakraborty, "Decision making for material selection using the UTA method", *Int J Adv Manuf Technol*, vol. 57, pp. 11–22., 2011.

[13] S. Chakraborty, E. K. ZAVADSKAS, "Applications of WASPAS Method in Manufacturing Decision Making", *Informatica*, vol. 25(1). pp. 1–20., 2014.

[14] S. J Chen, C. L Hwang, *Fuzzy Multiple Attribute Decision Making: Methods and Applications*. Springer-Verlag, Berlin, 1992.

[15] A. Jahan, F. Mustapha, M.Y. Ismail, S.M. Sapuan, M. Bahraminasab, "A comprehensive VIKOR method for material selection", *Materials and Design*, vol. 32, pp. 1215–1221, 2011.

[16] S. Opricovic, *Multicriteria Optimization of Civil Engineering Systems*, Faculty of Civil Engineering, Belgrade, 1998.

[17] E. K. Zavadskas, Z. Turskis, J. Antucheviciene. A. Zakarevicius, "Optimization of weighted aggregated sum product assessment", Electronics and Electrical Engineering, vol. 6(122), pp. 3-6., 2012.

[18] M. Farag, *Materials selection for engineering design,* New York: Prentice-Hall; 1997

Software Framework for REST Client Android Applications: Canvas LMS Case Study

Milan Pandurov, Srđan Milaković, Nikola Lukić, Goran Savić, Milan Segedinac, Zora Konjović
Faculty of Technical Sciences, University of Novi Sad
milanpandurov@gmail.com, srki@outlook.com, luknik94@gmail.com, {savicg, milansegedinac, ftn_zora}@uns.ac.rs

***Abstract* – The paper presents a software framework for developing Android applications. The framework has been designed for mobile applications which mainly operate as a thin client for accessing functionalities provided by RESTful web services. The framework contains four layers. The first layer provides network communication with RESTful services. The second one is responsible for the storage and processing of the data at the client side, while the third layer handles interaction with a user. The fourth layer serves as a mediator between data storage and user interface. The proposed framework has been verified on the case study of developing mobile Android application for Canvas LMS. The application communicates with REST API of Canvas LMS enabling users to use common Canvas features on a mobile device.**

INTRODUCTION

Modern software applications must deal with increasing diversity of software and hardware platforms, as well as with issues related to users' mobility. This has led us to current trends in developing internet-based applications where a server provides core functionalities that are accessed from (usually "thin") client applications. A traditional approach for accessing internet-based applications included internet browser which was installed on a personal computer. Wider use of mobile devices has set a demand to access server functionalities using native mobile applications. Such applications have been specifically written for an operating system executing on a mobile device. Currently, the most popular applications of this type are those written for Android and iOS operating systems. When it comes to internet-based applications, beside platform-dependent differences, all client applications access the same core functionalities which are set on server computers. Client applications vary only in the manner in which they communicate with the server and handle user interaction. For this reason, there have been developed platform agnostic mechanisms within server-side applications that can be accessed from any client platform.

Web services [1] are the standard solution for cooperation between various client platforms. Lately, a particularly popular implementation of web services is RESTful [2], which is based on REST software architecture [3].

RESTful services provide client with a uniform access to system resources. In REST terminology, a *resource* includes data or functionalities, where each resource is identified by its uniform resource identified (URI). A client interacts with RESTful services through very limited set of operations with predefined semantics. Typically supported operations are *create* (PUT), *read* (GET), *update* (POST) and *delete* (DELETE). Operation GET gets the current state of the resource. The purpose of the POST operation is to change resource's state. Operation PUT creates a resource, while DELETE operation removes it. Resources itself are separated from their representation format. It means that the same resource may be transferred in different formats, such as HTML, XML, JSON, etc. Since REST architecture proposes a stateless protocol for communication between client and server, RESTful services are designed to use HTTP protocol.

This paper presents the architecture of a software framework for developing client Android applications that access server functionalities which are exposed as RESTful web services. Since this type of Android applications is very common currently, the paper gives a general development framework neutral to any specific domain or application functionalities.

The proposed framework has been verified on developing an Android application for Canvas LMS [4]. The application access RESTful web services of Canvas LMS enabling students to access common information about courses they are enrolled.

RELATED WORK

Our paper is focused on client applications that meet following requirements:

- They receive/send data through network from/to RESTful web services contained within the server-side application
- They have its local data storage that cache data received from server
- They implement their own UI logic

Dobjanschi in [5] proposed software patterns for communication with RESTful services from Android application. The proposed patterns send/receive data using Android Service API or Android Content Provided API. The web page [6] contains a source code of sample implementation of this pattern.

Most solutions for communication between Android application and RESTful services are based on Dobjanschi's work. His solution is primarily focused on the communication with services, not considering other components in the Android application that should process, store and display data received from the network.

When it comes to data management in a client application, besides data fetching, the application must store some of the received data locally on the mobile device. This local cache may speed up the application and

optimize network traffic, given that application fetches data only when necessary. This implies data synchronization between client and server. The paper [7] proposes different synchronization patterns which can be used for this purpose. Next chapter describes synchronization methods used in the framework proposed in this paper.

UI layer of Android application must provide various visual controls for user interaction, while being at the same time flexible enough to combine these controls on the device display dynamically. A standard approach to provide this feature is by using Android Fragments [8]. A fragment [9] is a reusable UI component with UI controls and its own UI logic. It is displayed within an Android Activity element. A single fragment may contain its subfragments, while an activity may be composed of multiple fragments. A fragment can be dynamically added, removed or replaced within an activity.

Aforementioned solutions give proposals for individual requirements of REST-based Android applications. This paper's aim is to propose a comprehensive framework which covers all layers of Android applications of this type. The framework proposed in this paper implements common functionalities for each layer, while covering the connections between layers, too. The next chapter proposes the framework architecture.

PATTERN COMPONENTS

For many years, *de facto* standard for developing modular applications is a Model-View-Controller (MVC) software pattern. This pattern distinguishes three general application layers: data representation layer (Model), user interaction layer (View), and a layer that links data with displaying logic (Controller).

The framework proposed in this paper is also MVC-based. Figure 1 shows the general architecture of the framework and its place within REST-based Android application. An android application accesses server's REST API by executing REST methods. The server sends its response formatted as a JSON object. Android application is set on a mobile device and organized in conformance with our framework, which is MVC-based.

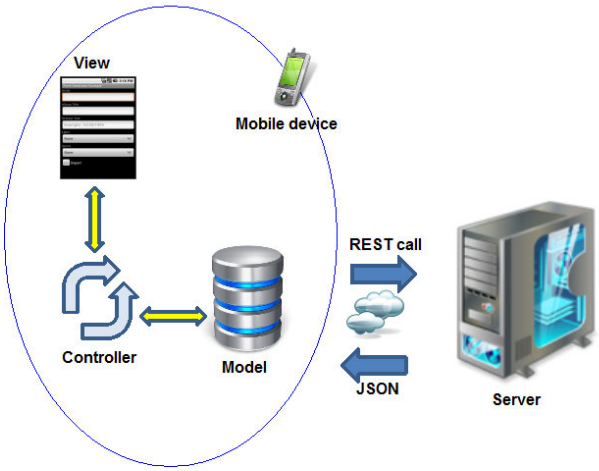


Figure 1. Framework global architecture

For each component in the framework, further text describes its functionalities and subcomponents.

As mentioned, *Model* component provides data management. Figure 2 shows its subcomponents.

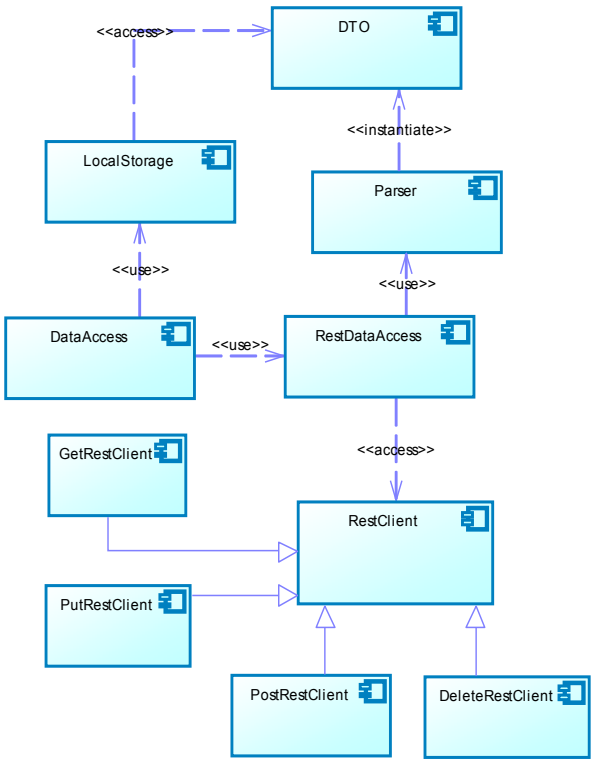


Figure 2. Model component

DataAccess is a proxy for data management. Other global components (like *Controller*) communicate with *DataAccess* to fetch/send data. Data is fetched from the application's local storage. For storing data on the mobile device, the framework uses *LocalStorage* component.

The in-memory representation of the data is given within *DTO* (*Data Access Object*) component. This component contains an object-model of the application's data, where each domain entity should be represented with an appropriate class holding entity's data.

When the application starts there is no application's data on the mobile device. They should be obtained from the server through its REST API. *RestDataAccess* component provides this functionality. The network communication with the server is performed within *RestClient* component. For the communication with REST services, the framework uses built-in Android classes that execute REST methods through HTTP protocol. With regard to the differences between REST methods, we propose subcomponents for executing *GET*, *PUT*, *POST* and *UPDATE* REST methods, respectively.

During the communication with the server, data will be passed in JSON format. *Parser* component is responsible for data conversion from JSON format to DTO in-memory representation, which is used by application local storage.

UI layer rely on built-in Android UI components. Figure 3 shows this layer architecture.

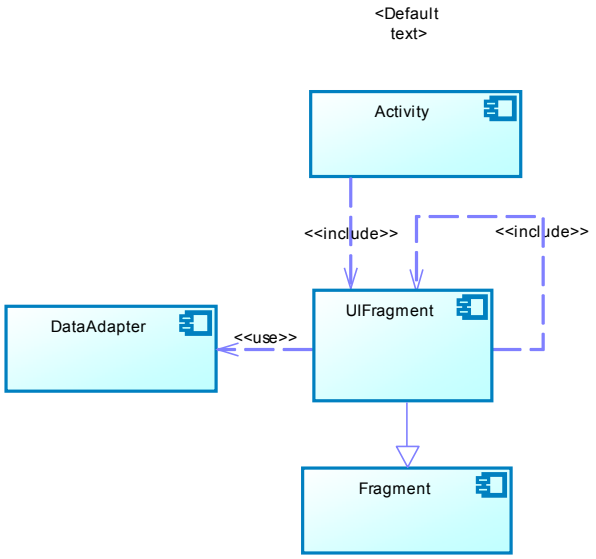


Figure 3. View component

The framework proposes just a single *Activity* element within a whole application. This *Activity* will contain different *Fragment* elements that provide interaction with a user. Fragments are represented by *UIFragment* component whose implementation relies on built-in Android *Fragment* class. *UIFragment* may contain child *UIFragment* elements. Fragment is populated with content using *DataAdapter* component. For each UI control, *DataAdapter* component will contain a separate class that contains specific logic for setting control's content. When a user navigates within the application, only fragments contained in the *Activity* element will be changed, while the Activity remains the same. Fragment management is responsibility of *Controller* layer, which is described below.

Controller layer manages UI navigation and links UI layer with data defined within *Model* layer. The architecture of *Controller* layer is shown in Figure 4.

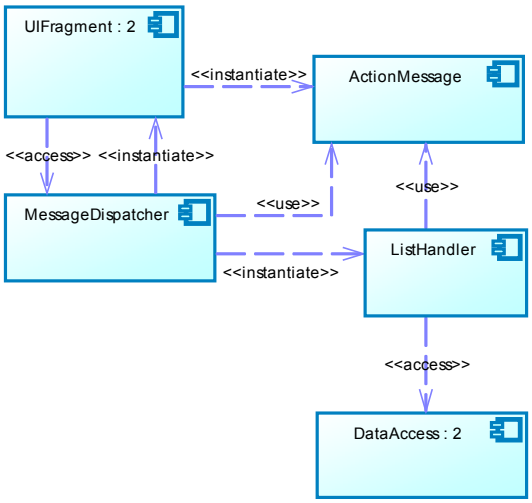


Figure 4. Controller component

UIFragment, as a part of user interface, reacts to the specific UI event (e.g. button-click) and generates a message represented by *ActionMessage* component. The message holds information about UI event. *UIFragment* passes the message to *MessageDispatcher*, which instantiates a new *UIFragment*, if necessary. This new fragment can replace current fragment in the *Activity*. *UIFragment* displays data received from *Controller* layer. *Controller* loads data from *Model* layer using *ListHandler* component which is a bridge between *Controller* and *DataAccess* component.

AUTHENTICATION

Concerning the fact that most of the REST-based server applications support user management, in addition to mention components, the framework introduces a special-purpose component providing user authentication. Since they are typically based upon the stateless HTTP protocol, the authorization in most of the REST-based server applications is achieved by using access tokens. An access token is a random server generated opaque string assigned to each authenticated user when logging in to the system. The user is required to introduce itself to the server by sending the access token along with every request. The main advantage of such an approach is the fact that clients never send their passwords (not even encrypted) to the server, but only random strings. In addition to that, it is possible to restrict the duration of an access token and to deactivate the access token manually, so to prevent possible abuses.

User authentication and access tokens generation in the proposed component is based upon OAuth2 protocol. Such an approach prevents client application to abuse the data entered when logging into the system, since entering usernames and passwords is delegated to the server, while client application manages only access tokens. OAuth2 protocol requires each client application to have its ID and secret key, obtained when registering the application to the system. Figure 5 shows the sequence of activities performed when authenticating a user in a system that uses OAuth2 protocol.

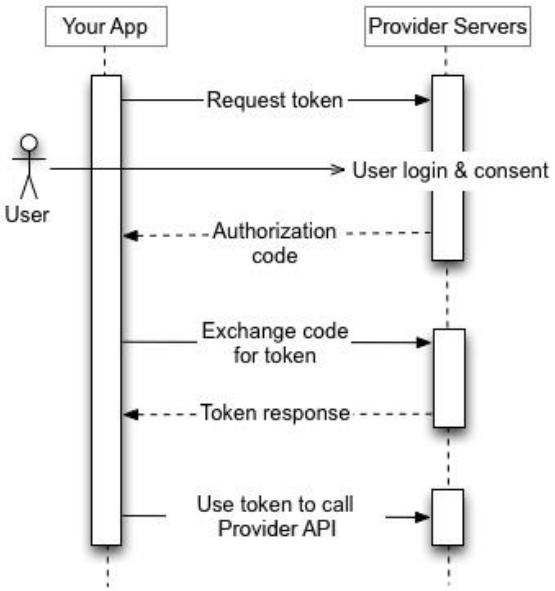


Figure 5. OAuth2 authentication

The first step in the authentication process consists of redirecting the user to the server application's page where he or she enters username and password. If the entered data are correct, the server application assigns temporary code to the user. After that, the client application sends the temporary code and the application's secret key to server

application. The response to this request contains an access token that is being sent along with every other request.

This authentication mechanism is suitable when client is a Web application so that the secret key needs not be persisted at the client's local machine. In cases when client is a mobile or a desktop application there is a risk of stealing secret keys by decompiling the executable code of the application (if the secret key is hardcoded into the application or if the application keeps the secret key in working memory).

To avoid this problem, a new component is introduced: a Web application that mediates in the process of assigning the access tokens, namely Authentication server. The sequence of authentication activities in such a system is shown in figure 6.

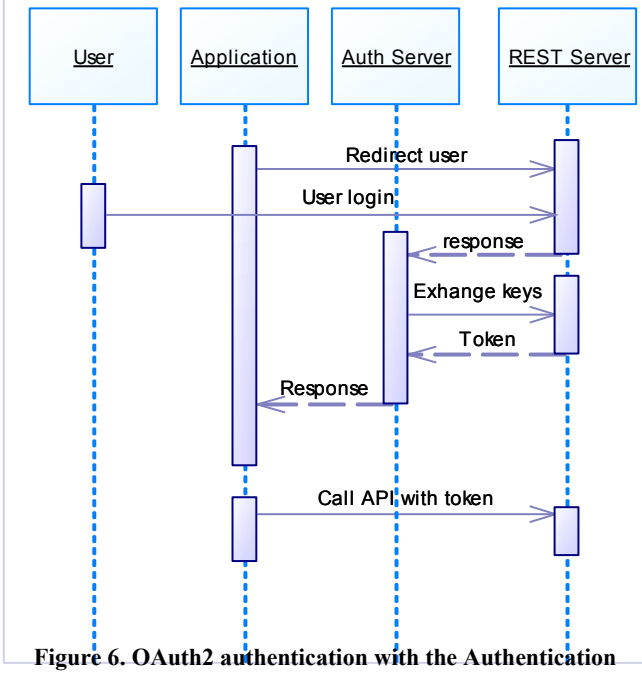


Figure 6. OAuth2 authentication with the Authentication server

In the proposed approach, instead of letting client application itself to keep the temporary code, the temporary code is being kept by the Authentication server, which holds the secret key as well. The Authentication server's task is to complete the authentication process and to send the access token to the client application. This mechanism denies client application ever to get hold of the secret key and reduces the risk to stealing the user's access token. Even if this would happen, the overall system integrity would not be threatened.

CANVAS ANDROID CLIENT

Canvas LMS [4] is an open source cloud-native learning management system developed by Instructure. This LMS provides a wide range of e-learning functionalities based upon web 2.0 technologies, such as e-learning tools, tools that support collaborative work and system administration tools [10]. Canvas LMS itself is a Ruby on Rails application, but it can be easily extended with third party tools regardless of the particular language in which the tools have been developed, since it supports IMS LTI standard.

Instructure offers a mobile application for Canvas [11]. Even though Canvas LMS is an open source application, Canvas for Android is not open source and it can be used only in combination with the Instructure hosted instances of Canvas LMS. Therefore, this paper proposes an open source Android application for Canvas LMS (Canvas Android Client - CAC) that can be used in educational settings with self-hosted instances of Canvas LMS. CAC is based upon the framework for developing android REST clients proposed in the previous parts of this paper.

Canvas LMS has an open REST API through which it exposes some of its operations as external services. The operations that are exposed via the API include, among others, mechanisms for accessing assignments, course information, registration, roles, users and discussion topics. Full specification of Canvas API can be accessed at [12]. CAC is a thin REST client, and its entire functionality comes down to calling the services from Canvas REST API and interpreting the results. Since calling the REST services requires the client to be authenticated, the authentication mechanism is described in details in the *Canvas authentication* section that immediately follows. The set of functionalities that are supported by the current version of CAC is described in the *CAC functionalities* section.

The current version of CAC supports following functionalities:

- Browsing courses
- Browsing announcements
- Browsing assignments

All the functionalities require user to be authenticated. After the user has successfully logged in to the system, she or he can choose the action from the menu shown in figure 7.

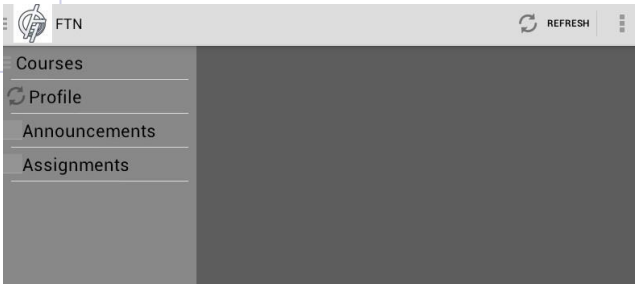


Figure 7. CAC main menu

When a user chooses the item *Courses* from the main menu, the list of all courses is being presented, as shown in Figure 8. Each item in this list has course name and the information that indicates if the course is still available.

When user chooses one item from the list, a view with the details about the selected course is being shown, as in the figure 9. In addition to the general information on the course, this view contains a list of announcements and a list of assignments from the selected course.

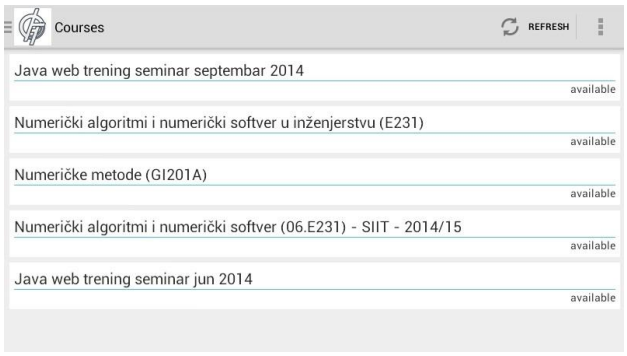


Figure 8. Courses

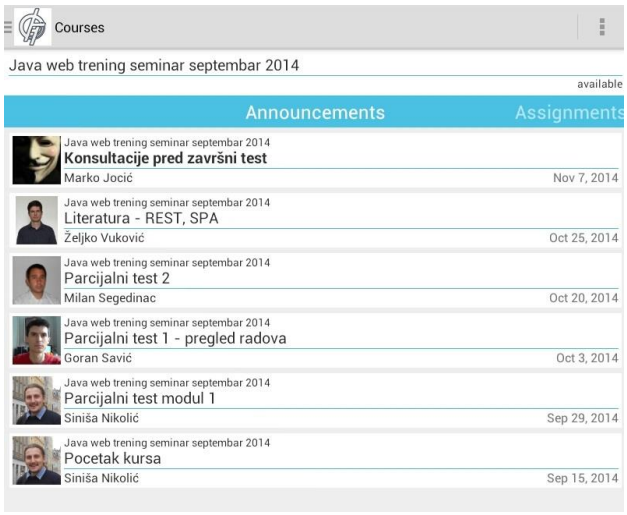


Figure 9. Course details

When the user chooses the item Announcements from the main menu (figure 7), the list of all announcements from all the courses that the user is enrolled is being presented, as shown in figure 10. Each item in this list contains the avatar of the user who has posted the announcement, course name, as well as the announce title and date the announcement has been posted.

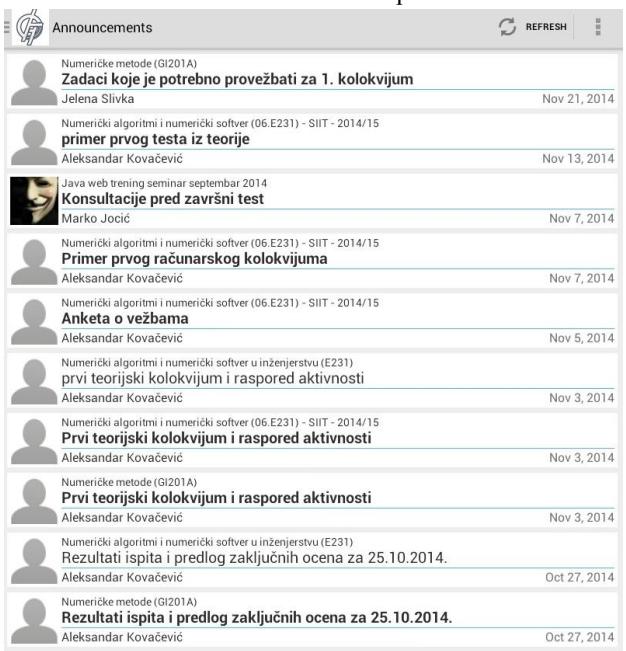


Figure 10. Announcements

When an item in the list is selected, the general information along with the content of the announcement is shown, as presented in figure 11.

Figure 11. Announcement details

When a user chooses the item assignments from the main menu (figure 7), the list of all current assignments from all the courses that the user is enrolled part in is being shown, as presented in figure 12. Each item in this list contains course title, assignment name as well as the date upon which the assignment is available.

Figure 12. Assignments

When one item from the assignments list is being selected, a view with the assignment details containing assignment general information along with the content of the assignment is being presented, as shown in figure 13.

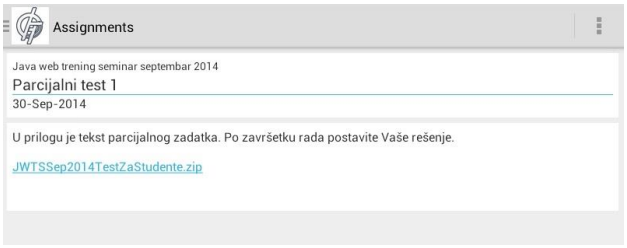


Figure 13. Assignment details

CONCLUSION

A new software framework for developing REST-based Android applications has been proposed. The framework can be used for any thin-client Android application that access functionalities exposed as REST web services. For this type of Android applications, the framework proposes application architecture, as well as particular implementation of common functionalities. The framework generally follows MVC pattern, organizing programming code into three layers where each layer provides a single point for communication with other layers. Concerning the fact that most of the REST-based server applications support user management, in addition to mention components, the framework introduces a special-purpose component providing user authentication.

The framework has been evaluated on the Android application for Canvas LMS. In contrast to the official Canvas for Android application developed by Instructure, which can be used only in combination with Instructure-hosted Canvas instances, the application proposed in this paper is open-source and can be used with self-hosted Canvas instances. Current version of the application provides access to common Canvas features within mobile environment.

The future plans for the proposed application include:

- Start using the application at the Faculty of Technical Sciences at University of Novi Sad
- Extending current set of the application's functionalities with new features
- Integrating the application with other educational services at University of Novi Sad

ACKNOWLEDGMENT

Results presented in this paper are part of the research conducted within the Grant No. III-47003, Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] W3C (2004), Web Services Glossary". Retrieved 24.11.2014

[2] L. Richardson, S. Ruby (2007), RESTful web service, O'Reilly Media, ISBN 978-0-596-52926-0

[3] R. Fielding, R. Taylor (2002), Principled Design of the Modern Web Architecture (PDF), *ACM Transactions on Internet Technology* (*TOIT*) (New York: Association for Computing Machinery) 2 (2): 115–150, doi:10.1145/514183.514185, ISSN 1533-5399

[4] Instructure Inc. (2012), Canvas LMS, http://www.instructure.com, Retrieved: 24.11.2014.

[5] V. Dobjanschi (2010), Developing Android REST Client Applications, https://dl.google.com/googleio/2010/android-developing-RESTful-android-apps.pdf, Retrieved 24.11.2014.

[6] CodeProject (2012), Sample Implementation of Virgil Dobjanschi's Rest pattern, http://www.codeproject.com/Articles/429997/Sample-Implementation-of-Virgil-Dobjanschis-Rest-p. Retrieved 24.11.2014.

[7] Z. McCormick and D. C. Schmidt (2012), Data Synchronization Patterns in Mobile Application Design, *Proceedings of the Pattern Languages of Programs* (*PLoP*) *2012 conference*, Tucson, Arizona.

[8] J. Wilson (2013), Creating Dynamic UI with Android Fragments, *Packt publishing*, ISBN: 9781783283095

[9] Android Developers (2014), Fragments, http://developer.android.com/ guide/components/fragments.html. Retrieved 24.11.2014.

[10] N. Nikolić, G. Savić, M. Segedinac and Z. Konjović (2014), Migration from Sakai to Canvas, P*roceedings of the 4th International Conference on Information Society and Technology (ICIST 2014)*, Kopaonik, Serbia, 366 – 370, ISBN: 978-86-85525-14-8

[11] Instructure Inc (2014), Canvas for Android, https://play.google.com/store/apps/details?id=com.instructure.candroid, Retrieved: 26.11.2014.

[12] Instructure Inc. (2012), Canvas LMS API Documentation, https://canvas.instructure.com/doc/api/, Retrieved: 26.11.2014.

Bioinspired metaheuristic algorithms for global optimization

Marko Mitić*, Najdan Vuković**, Milica Petrović*, Jelena Petronijević*, Ali Diryag*, Zoran Miljković*
* University of Belgrade - Faculty of Mechanical Engineering/Production Engineering Department, Belgrade, Serbia
** University of Belgrade - Faculty of Mechanical Engineering/Innovation Center, Belgrade, Serbia
mmitic@mas.bg.ac.rs, nvukovic@mas.bg.ac.rs, mmpetrovic@mas.bg.ac.rs, jpetronijevic@mas.bg.ac.rs, ali6981@gmail.com, zmiljkovic@mas.bg.ac.rs

***Abstract*—This paper presents concise comparison study of newly developed bioinspired algorithms for global optimization problems. Three different metaheuristic techniques, namely Accelerated Particle Swarm Optimization (APSO), Firefly Algorithm (FA), and Grey Wolf Optimizer (GWO) are investigated and implemented in Matlab environment. These methods are compared on four unimodal and multimodal nonlinear functions in order to find global optimum values. Computational results indicate that GWO outperforms other intelligent techniques, and that all aforementioned algorithms can be successfully used for optimization of continuous functions.**

I. INTRODUCTION

In recent years, various nonlinear optimization problems are solved using biologically inspired solutions. Main reason lies in a fact that, in these cases, traditional algorithms often fail in producing wanted/expected results. Bioinspired metaheuristic techniques represent a well-known mathematical tool for solving hard optimization tasks that cannot be solved using other approaches. This study represent a concise comparison of three such algorithms for finding global optimum of continuous nonlinear functions.

Bioinspired metaheuristic algorithms mimic the behaviour of animals in nature by turning their swarming, flocking or grouping into mathematical procedures. In these intelligent methods, an algorithm starts with random population of individuals that are next grouped around optimal solution using iterative search. In comparison with single-based algorithms, population-based metaheuristic has significant advantages in finding overall best optimization result [1]:

- Multiple candidate solutions share information about the search space which results in sudden jumps toward the promising part of search space.
- Multiple candidate solutions assist each other to avoid locally optimal solutions.
- Population-based meta-heuristics generally have greater exploration compared to single solution-based algorithms.

These clear advantages influence the development of large number of new bioinspired optimization algorithms over the last decade. Most popular techniques in the field include, firefly algorithm [2], cuckoo search [3], bat algorithm [4], grey wolf optimizer [1], and particle swarm optimization [5], which are successfully applied for solving various engineering problems [6,7,8,9,10]. These studies prove that metaheuristic algorithms are superior in avoiding stagnation in local solution due to their stochastic nature

Despite the aforementioned, the main disadvantage of metaheuristic methods lies in the fact that there is no guarantee that found solution is actually the optimal one. Likewise, in some cases the algorithm dependent parameters is hard to determine. However, in most real world problems, the search space is usually unknown and prone to large number of local optimums, so the metaheuristics with the ability of extensive search in this space represents good option.

In this paper, three popular metaheuristic techniques, namely Accelerated Particle Swarm Optimization (APSO), Firefly Algorithm (FA), and Grey Wolf Optimizer (GWO) are compared in optimization task of different nonlinear functions. These algorithms are tested on four unimodal and multimodal well-known benchmark problems. Results show the efficiency of bioinspired methods since in all cases, the experimentally obtained best algorithm successfully converged.

The paper is organized as follows. After the brief introduction, the main optimization methods are introduced in the second section. In the third part of the paper mathematical description of nonlinear functions is presented. Fourth section show the obtained computational results. Finally, the last section gives the overall conclusion of this study.

II. BIOINSPIRED ALGORITHMS

In this section mathematical description of each intelligent method is given in brief.

A. *Accelerated Particle Swarm Optimization - APSO*

This algorithm is inspired by fish schooling behavior in nature [5,10]. Each individual (e.g. particle) in swarm flies toward its best and currently best solution of the given problem. Likewise, the algorithm incorporates random component, so the global search is to some point stochastic. Two main equations of the traditional particle swarm optimization algorithm are [10, 11]:

$$v_i^{t+1} = v_i^t + \alpha r_1 \odot \left[g^* - x_i^t \right] + \beta r_2 \odot \left[x_i^* - x_i^t \right] \quad (1)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (2)$$

where: symbol $\odot$ is the Hadamard product, v_i^t and x_i^t are current velocity and position of the particle, respectively, v_i^{t+1} and x_i^{t+1} are next velocity and position of an individual respectively, r_1 and r_2 are two random vectors, α and β are algorithm's learning parameters. In recent research work, an accelerated version of this algorithm is introduced [11]. Velocity in APSO is calculated with

$$v_i^{t+1} = v_i^t + \alpha r(t) + \beta\left[g^* - x_i^t\right], \tag{3}$$

where: g^* is global best solution, and r is drawn from standard Gaussian distribution. Equation (2) is also modified, so that location of the particle in APSO is calculated as

$$x_i^{t+1} = (1-\beta)x_i^t + \beta g^* + \alpha r\,. \tag{4}$$

Comparing (2) and (4) one can conclude that APSO does not include velocity parameter. APSO uses only parameters α and β, so it is much easier to initialize and to understand. These are the main reasons for the implementation of APSO over PSO in this research work.

B. Firefly Algorithm - FA

Social behavior of fireflies in nature served as an inspiration for developing FA [2].Two important issues of this method should be noted: the variation of light intensity in real fireflies and the formulation of attractiveness [12]. The light intensity parameter is of crucial importance since it represent fitness function (e.g. optimization goal), and influence the movement of the entire swarm. The attractiveness of one firefly to another individual in swarm corresponds to light intensity and is calculated as:

$$\beta = \beta_0 \cdot e^{-\gamma r^2}\,, \tag{5}$$

where: r is the Euclidian distance between two individuals, γ is the parameter of light absorption, and β_0 is the attractiveness at $r = 0$. The movement of firefly *i* towards firefly *j* is now defined as:

$$x_i = x_i + \beta_o e^{-\gamma r^2}(x_j - x_i) + \alpha\varepsilon_i\,, \tag{6}$$

where: ε_i is random Gaussian number, and α is the randomization parameter chose by the designer. Using this last two equations, (5) and (6), fireflies can be sorted in accordance with their light intensity (i.e. achieved performance), and then directed towards the better solution.

C. Grey wolf optimizer - GWO

It is known that grey wolves are considered top of the food chain, and are usually grouped in small packs. Of particular interest for the GWO method is that they have a very strict social dominant hierarchy [1]. The solutions in GWO are generated using individuals defined as: hierarchy leader (i.e. alpha), subordinate wolf in the second level (i.e. beta), third level individual (i.e. delta), and low-level wolf (i.e. omega).

Logically, the fittest GWO solution of the optimization problem is described by alpha. Similarly to this, second and third best solutions are beta and delta, respectively. Rest of the solutions are defined using omega wolves. Optimization with GWO is guided by alpha, beta and gamma solutions, obtained through processes of tracking, encircling and attacking prey. Mathematical model for encircling behavior is given with [1]

$$\vec{D} = \left|\vec{C}\cdot\vec{X}_p(t) - \vec{X}(t)\right| \tag{7}$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A}\cdot\vec{D} \tag{8}$$

where: $\vec{X}_p$ and $\vec{X}$ are position vector of a prey and grey wolf, respectively, $\vec{A}$ and $\vec{C}$ indicate coefficient vectors, and t is the current iteration. Coefficients are calculated with

$$\vec{A} = 2\vec{a}\cdot\vec{r}_1 - \vec{a}\,, \tag{9}$$

$$\vec{C} = 2\cdot\vec{r}_2\,, \tag{10}$$

where: $\vec{a}$ is linearly decreased over iterations from 2 to 0, and $\vec{r}_1$, $\vec{r}_2$ are random vectors chosen in the domain $[0,1]$. It is presumed that, for the hunting phase, alpha, beta and gamma have better knowledge about the location of the pray [1]. Therefore, all other individuals (i.e. omega wolves) update their position based on the information of these aforementioned three best solutions. Equations that describe this most important step are as follows

$$\begin{aligned}\vec{D}_\alpha &= \left|\vec{C}_1\cdot\vec{X}_\alpha - \vec{X}\right|,\\ \vec{D}_\beta &= \left|\vec{C}_2\cdot\vec{X}_\beta - \vec{X}\right|,\\ \vec{D}_\delta &= \left|\vec{C}_3\cdot\vec{X}_\delta - \vec{X}\right|,\end{aligned} \tag{11}$$

$$\begin{aligned}\vec{X}_1 &= \vec{X}_\alpha - \vec{A}_1\cdot\left(\vec{D}_\alpha\right),\\ \vec{X}_2 &= \vec{X}_\beta - \vec{A}_2\cdot\left(\vec{D}_\beta\right),\\ \vec{X}_3 &= \vec{X}_\delta - \vec{A}_3\cdot\left(\vec{D}_\delta\right),\end{aligned} \tag{12}$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3}\,. \tag{13}$$

Vector $\vec{A}$ is responsible for attacking the prey, i.e. for exploitation phase. The value $\vec{a}$ is linearly decreasing, and therefore is parameter $\vec{A}$. It is shown that $|A| < 1$

forces the wolves to attack towards the pray [1]. Unlike exploitation vector, parameter $\vec{C}$ favors the exploration. This component provides random weights for prey in order to stochastically emphasize ($C > 1$) or deemphasize ($C < 1$) the effect of prey in defining the distance [1]. More information on the algorithmic procedure of GWO one can find in [1].

III. BENCHMARK PROBLEMS

We tested these described algorithms on four unimodal and multimodal nonlinear functions. They are chosen in order to reflect different sorts of real world optimization problems. The goal in each case is to determine optimum value of function, which varies depending on the chosen function. The mathematical description of the functions is given in Table I.

TABLE I.
MATHEMATICAL DESCRIPTION OF NONLINEAR FUNCTIONS

Function ID	Mathematical description	Type
F1	$f_1(x) = \sum_{i=1}^{n} x_i^2$	Unimodal
F2	$f_3(x) = -\cos(x)\cdot\cos(y)\cdot e^{\left[-\left((x-\pi)^2+(y-\pi)^2\right)\right]}$	Unimodal
F3	$f_2(x) = \sum_{i=1}^{n-1}\left[(x_i-1)^2 + 100\left(x_{i+1}-x_i^2\right)^2\right]$	Multimodal
F4	$f_4(x) = -\sum_{i=1}^{n}\sin(x_i)\cdot\left[\sin\left(\frac{ix_i^2}{\pi}\right)\right]^{20}$.	Multimodal

Visual representations of these functions are given on Fig. 1 - Fig. 4. One can note the nonlinear nature of these functions, which makes them a fairly demanding problems for optimization.

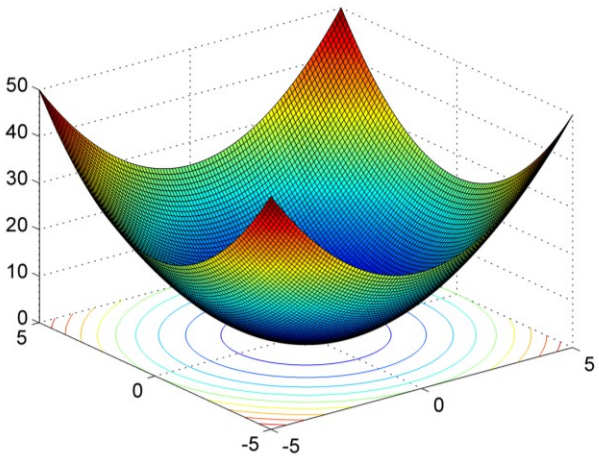


Figure 1. Sphere function (F1) in two dimensions.

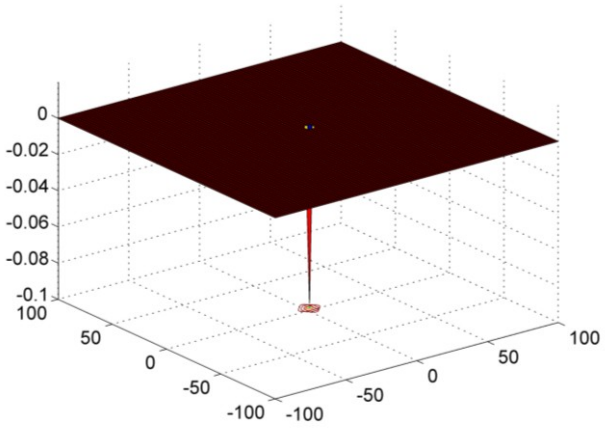


Figure 2. Easom function (F2) in two dimensions.

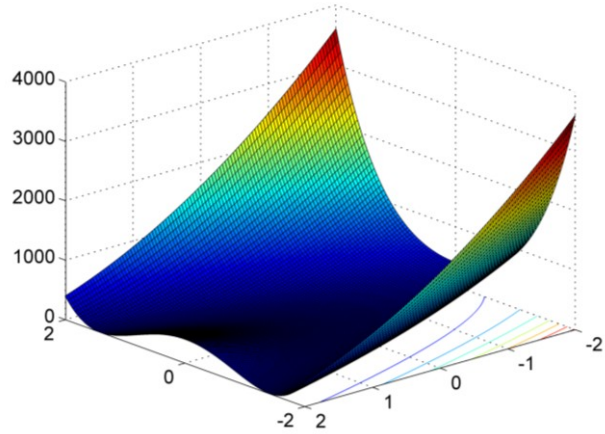


Figure 3. Rosenbrock function (F3) in two dimensions.

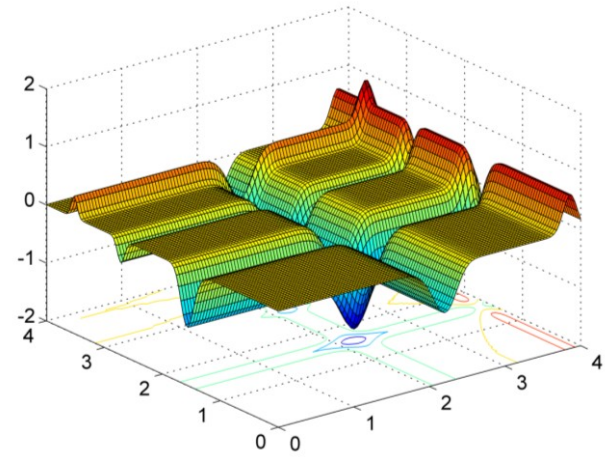


Figure 4. Michalewicz function (F4) in two dimensions.

IV. COMPUTATIONAL RESULTS

Experimental results for each described algorithm are given in this section. Reported results is derived using Matlab software using laptop PC with 4GB of RAM that runs on 64-bit Windows 7. Maximum number of iterations and algorithms specific parameters are chosen to be recommended values. Each algorithm is tested 30

times for each optimization problem in order to obtain a statistical evaluation of method's performance. The initial population is scattered across the search space at the beginning of each population. The algorithm dependent parameters are given in Table II. Fig. 5 - Fig. 8 show top view of these functions with swarm individuals dispersed across the entire surface. Finally, Table III - Table V present the computational results, in which best result refers to the determined optimum value of a particular metaheuristic algorithm for a given function.

TABLE II.
ALGORITHM DEPENDENT PARAMETERS

Algorithm	Specific parameter values
APSO	$\beta = 0.5$; $\alpha = 0.7^i$, i=*num_generation*
FA	$\beta_0 = 1$; $\gamma = 1$; $\alpha = 0.2$
GWO	$a = 2 - iter \cdot ((2)/Max\_iter)$ r_1, r_2 - *random numbers in* $[0,1]$

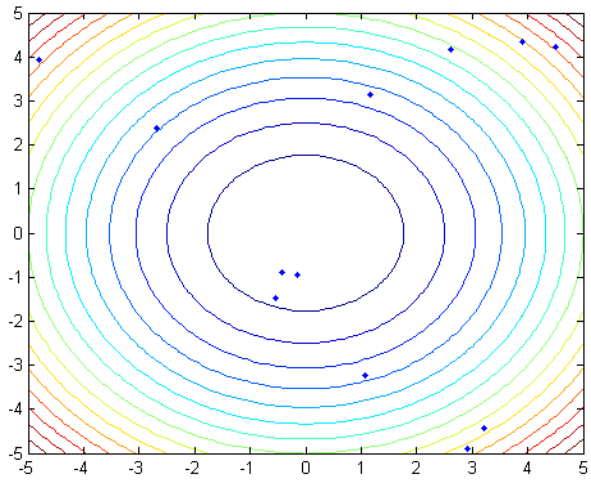


Figure 5. Initial population over the surface of Sphere function (top view).

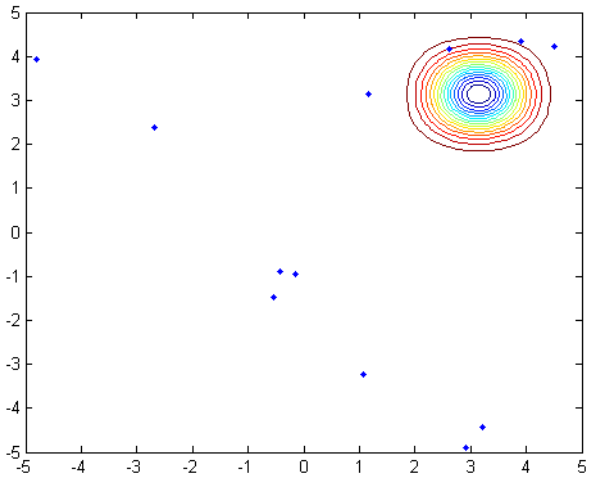


Figure 6. Initial population over the surface of Easom function (top view).

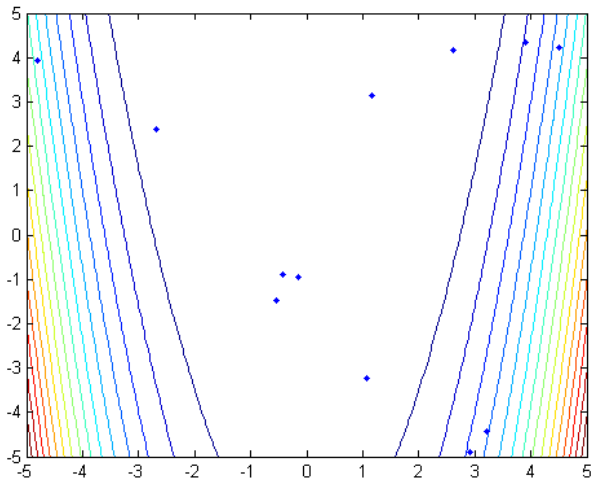


Figure 7. Initial population over the surface of Rosenbrock function (top view).

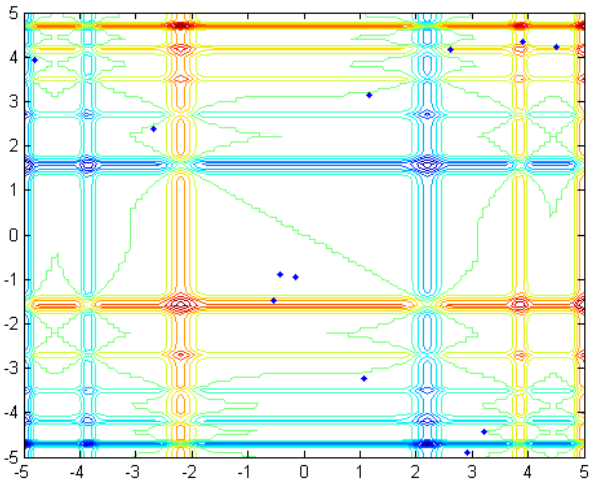


Figure 8. Initial population over the surface of Michalewicz function (top view).

TABLE III.
COMPUTATIONAL RESULTS FOR ACCELERATED PARTICLE SWARM OPTIMIZATION

Function ID	Best result	Mean value	Convergence success rate
F1	0	$6.8777*10^{-6}$	100%
F2	-1	-1	100%
F3	$-1.7909*10^{-5}$	0.1339	63.33%
F4	-1.8011	-1.7547	90%

TABLE IV.
COMPUTATIONAL RESULTS FOR FIREFLY ALGORITHM

Function ID	Best result	Mean value	Convergence success rate
F1	0	0.0419	100%
F2	-0.9748	-0.9748	100%
F3	0.5913	0.5913	0%
F4	-1	-1	0%

TABLE V.
COMPUTATIONAL RESULTS FOR GREY WOLF OPTIMIZER

Function ID	Best result	Mean value	Convergence success rate
F1	0	10^{-10}	100%
F2	-1	-0.9999	100%
F3	0	0	100%
F4	-1.8013	-1.8012	100%

Comparing Table III, Table IV and Table V one can conclude that GWO algorithm obtained best results. Convergence rate also indicate the superiority of GWO in comparison with other methods. The convergence is determined experimentally using ""convergence to an optimum criteria" [11], which indicate that the algorithm successfully converges when its best solution reaches ε likelihood of theoretical optimum for a given problem. Therefore, it can be concluded that GWO is the most reliable of all tested metaheuristic methods.

GWO algorithm also show optimal performance in terms of best obtained results and also mean value over 30 independent runs. This corresponds to some other optimization studies that indicate the same conclusion [1]. Second best result show APSO, while FA gives overall worst result. However, it should be noted that these problems are fairly hard to solve, and that both APSO and FA show better performance in optimizing real engineering problems [10,11,12,13,14].

Future research will include more bioinspired algorithms which will be tested on larger number of functions. Likewise, comparison of these intelligence methods should be studied on real world engineering problems; for example, metaheuristic methods can be employed for intelligent transportation in indoor environment using nonholonomic differential drive mobile robots [6].

V. CONCLUSIONS

In this paper, a concise experimental comparison study on bioinspired algorithms on function optimization is given. The study includes three newly developed metaheuristic techniques, namely Accelerated Particle Swarm Optimization (APSO), Firefly Algorithm (FA), and Grey Wolf Optimizer (GWO). These intelligent techniques are tested in optimization task of four different unimodal and multimodal functions. Each algorithm is tested 30 times on each optimization problem in order to obtain statistical evaluation of the experiments. Computational results show that GWO algorithm is the best one, and that it successfully converge in each simulation run with rate of 100% for all optimization tasks.

ACKNOWLEDGMENT

This work is supported by the Serbian Government - the Ministry of Education, Science and Technological Development - Project title: An innovative, ecologically based approach to the implementation of intelligent manufacturing systems for the production of sheet metal parts (2011-2015) under grant TR35004.

REFERENCES

[1] S. Mirjalili, S.M. Mirjalili and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Soft.*, vol. 69, pp. 46-61, 2014.

[2] X. S. Yang, "Firefly algorithms for multimodal optimization," in: *Stochastic algorithms: foundations and applications*, Springer Berlin Heidelberg, pp. 169-178, 2009.

[3] X. S. Yang, and S. Deb, "Engineering optimisation by cuckoo search," *Int. J. Math. Model. Num. Opt.*, vol. 1, pp. 330-343, 2010.

[4] X. S. Yang, "A new metaheuristic bat-inspired algorithm," in: *Nature inspired cooperative strategies for optimization* (NICSO 2010), pp. 65-74, 2010.

[5] J. Kennedy and R. Eberhart, "Particle swarm optimization," in: *IEEE/RSJ International Conference on Neural Networks*, pp. 1942-1948, 1995.

[6] M. Mitić and Z. Miljković, "Bio-inspired approach to learning robot motion trajectories and visual control commands," *Expert Syst. Appl.*, vol. 42, pp. 2624 - 2637, 2015.

[7] G. Wang, L. Guo, H. Duan, L. Liu and H. Wang, "A bat algorithm with mutation for UCAV path planning," *Sci. World J.*, pp. 418946, 2014.

[8] S. Karthikeyan, P. Asokan and S. Nickolas, "A hybrid discrete firefly algorithm for multi-objective flexible job shop scheduling problem with limited resource constraints," *Int. J. Adv. Manuf. Tech.,* vol. 72, pp. 1567-1579, 2014.

[9] M. R. Soltanpour and M. H. Khooban, "A particle swarm optimization approach for fuzzy sliding mode control for tracking the robot manipulator," *Nonlinear Dynam.*, vol. 74, pp. 467-478, 2013.

[10] A. H. Gandomi, G. J. Yun, X. S. Yang and S. Talatahari, "Chaos-enhanced accelerated particle swarm algorithm," *Commun. Nonlinear Sci. Numer. Simulat.*, vol. 18, pp. 327–340, 2013.

[11] X. S. Yang, "Engineering optimization: an introduction with metaheuristic applications," *John Wiley & Sons*, 2010.

[12] I. Fister, I. Jr. Fister, X. S. Yang and J. Brest, "A comprehensive review of firefly algorithms," *Swarm Evol. Comput.*, vol. 13. pp. 34-46, 2013.

[13] A. H. Gandomi, X. S. Yang, S. Talatahari and A. H. Alavi, "Firefly algorithm with chaos," *Commun. Nonlinear Sci. Numer. Simulat.* vol., 18, pp. 89–98, 2013.

[14] X. S. Yang, "Nature-inspired metaheuristic algorithms," *Luniver Press*, 2008.

Measuring influence of Facebook pages

Marko Jocić*, Đorđe Obradović*, Zora Konjović*
* University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia
m.jocic@uns.ac.rs, obrad@uns.ac.rs, ftn_zora@uns.ac.rs

Abstract — In this paper we propose a method for determining measure of influence of Facebook pages. Three characteristics of a Facebook page are measured: total number of fans of Facebook page, activity of Facebook page (frequency of posting) and post engagement of fans (weighted sum of number of likes, comments and shares) per post. Fuzzy approach is used to describe and determine measure of influence of Facebook pages. Finally, the paper presents results (influential and non-influential Facebook pages refined based on page category) obtained by combining these three measures and querying the 1.3 million Facebook pages large database.

I. INTRODUCTION

The explosive growth of social media has provided millions of people the opportunity to create and share content on a scale barely imaginable a few years ago. Massive participation in these social networks is reflected in the countless number of opinions, news and product reviews that are constantly posted and discussed in social sites such as Facebook, Twitter, Instagram, Pinterest and more [1]. Ideas, opinions, and products compete with all other content for the scarce attention of the user community. In spite of the seemingly chaotic fashion with which all these interactions take place, certain topics manage to get an inordinate amount of attention, thus bubbling to the top in terms of popularity and contributing to new trends and to the public agenda of the community. One aspect is the popularity and status of given members of these social networks, which is measured by the level of attention they receive in the form of followers who create links to their accounts to automatically receive the content they generate. The other aspect is the influence that these individuals wield, which is determined by the actual propagation of their content through the network. This influence is determined by many factors, such as the novelty and resonance of their messages with those of their followers and the quality and frequency of the content they generate. Equally important is the passivity of members of the network which provides a barrier to propagation that is often hard to overcome. Thus gaining knowledge of the identity of influential and least passive people in a network can be extremely useful from the perspectives of viral marketing, propagating one's point of view, as well as setting which topics dominate the public agenda [1].

Influence has long been studied in the fields of sociology, communication, marketing, and political science. The notion of influence plays a vital role in how businesses operate and how a society functions—for instance, see observations on how fashion spreads and how people vote. Studying influence patterns can help us better understand why certain trends or innovations are adopted faster than others and how we could help advertisers and marketers design more effective campaigns [2].

In this paper, focus is on Facebook, because it is world's largest social network and has available application programming interface, or an API - Graph API [3], through which various data from Facebook can be fetched and later analyzed. Also, this research is focused on measuring influence of Facebook pages, which are public profiles specifically created for businesses, brands, celebrities, causes, and other organizations. Unlike personal profiles, pages do not gain "friends," but "fans" - which are people who choose to "like" a page. Facebook pages can gain an unlimited number of fans, differing from personal profiles, which has had a 5,000 friend maximum put on it by Facebook. Pages work similarly to profiles, updating fans with things such as statuses, links, events, photos and videos. This information appears on the page itself, as well as in its fans' personal news feeds.

This paper is laid out in the following chapters; the first chapter gives an introduction into the research conducted in this paper, as well as motivation for the research. Chapter II gives an overview of some related work that has been identified both in academic paper and practical both commercial and academic solution terms. Chapter III outlines some basic facts about fuzzy sets theory and Graph API, while Chapter IV describes the method to determine measure of influence of Facebook pages. Chapter V shows the results of the proposed method, concluding with directions of future research.

II. RELATED WORK

As noted, this section deals with existing methods for determining influence of social media users. Because this topic is very popular and profitable nowadays, there is plethora of commercial solutions available online concerning social mining and social media analysis.

Socialbakers [4] tracks, analyzes, and benchmarks over 8 million social profiles across all the major social platforms including Facebook, Twitter, YouTube, LinkedIn, Instagram, Google+ and VK. They have statistics that are free and available to everyone with daily updates and historical data up to 3 months. Included in these free statistics are Facebook pages, which can be filtered by page category (brands, celebrities, sports, etc.) and country, where users can view pages with largest audience (number of fans) and find fastest-growing pages in the last day, week or month. However, paid version of this service allows updates several times per day, reporting (executive and custom reports), reports exporting into several different formats, historical data up to 5 years, and many more premium content like finding key influencers, determining engagement rating and key performance indicators, etc.

Simply Measured [5] is the leading social media analytics platform, providing complete measurement and reporting for serious marketers in all major social platforms including Facebook, Twitter, Google+, Instagram, YouTube, Vine, LinkedIn, Tumblr. This service delivers profile analytics and audience insights, cross-channel analysis, content and campaign performance measurements, brand and hashtags monitoring, social advertising analytics, influence and sentiment analysis, and many more. Simply Measured provides insightful reports that identify the success of a campaign, which are given mainly through different charts, e.g. total engagement on page posts per day, top keywords within post comments, engagement on page post per post type (status, link, photo, video), top times and days for comments and many more. Unfortunately, this service is available only in paid version, and only samples of data are shown for free.

Trackur [6] allows full monitoring of all mainstream social media including Twitter, Facebook and Google+, but also news, blogs, reviews and forums. This service delivers executive insights including trends, keyword discovery, automated sentiment analysis and influence scoring. Trackur is mainly used for tracking certain brand's status – who, where and in what context is talking about it. Results are automatically scored positive, negative, or neutral, and they also show the influence of each person discussing a brand. All of these results are delivered almost in real-time, as many sources are updated every 30 minutes. Trackur is also a paid service, and there is no data available for free.

Klout [7] uses Twitter, Facebook, LinkedIn, Wikipedia, Instagram, Bing, Google+, Tumblr, Foursquare, YouTube, Blogger, WordPress, Last.fm, Yammer and Flickr data to create Klout user profiles that are assigned a "Klout Score". Klout scores range from 1 to 100, with higher scores corresponding to a higher ranking of the breadth and strength of one's online social influence. Klout suggest it's users shareable content that his/hers audience hasn't seen yet, and also tracks how retweets, likes and shares change user's Klout score. In order to get his/hers Klout score calculated, user must log in to Klout with a Facebook and/or Twitter account, and later to connect all other social media accounts – the more accounts are connected, the more precise and relevant Klout score is. This service also offers possibility of connecting influencers with brands, so that brands can hire relevant influencers for a certain marketing campaign. Klout is free for regular users, i.e. influencers, but paid for business users, i.e. brands and marketers.

As shown, services that offer insightful social media analysis are mostly paid, which is expected due to the importance and value of the data. Also, none of the mentioned services show their algorithms and methods used to infer influence of a certain individual; they mostly just outline these methods, but the details are kept secret. This caused many researches to develop their own algorithms if they can't afford such premium services.

III. PRELIMINARIES

Approach proposed in this paper consists of using fuzzy sets theory to describe and determine measure of influence of Facebook pages. Our previous work in mathematical models for describing imprecise data [8][9][10][11][12] was our main encouragement for using this fuzzy approach.

A. Fuzzy sets and fuzzy logic

Fuzzy set theory was formalised by Professor Lotfi Zadeh at the University of California in 1965 [13]. Fuzzy logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth-truth values between "completely true" and "completely false" thus enabling modes of human reasoning which are mostly approximate rather than exact.

The essential characteristics of fuzzy logic as founded by Lotfi Zadeh are as follows:

- In fuzzy logic, exact reasoning is viewed as a limiting case of approximate reasoning.
- In fuzzy logic everything is a matter of degree.
- Any logical system can be fuzzified.
- In fuzzy logic, knowledge is interpreted as a collection of elastic or, equivalently, fuzzy constraint on a collection of variables.
- Inference is viewed as a process of propagation of elastic constraints.

The definition of a fuzzy set then, from Zadeh's paper is:

Definition. Let X be a space of points, with a generic element of X denoted by x. Thus $X = \{x\}$**.** A fuzzy set A in X is characterized by a **membership function** $f_A(x)$ which associates with each point in X a real number in the interval $[0,1]$, with the values of $f_A(x)$ at x representing the "grade of membership" of x in A. Thus, the nearer the value of $f_A(x)$ to unity, the higher the grade of membership of x in A. [13]

Membership functions for fuzzy sets can be defined in any number of ways as long as they follow the rules of the definition of a fuzzy set. The shape of the membership function used defines the fuzzy set and so the decision on which type to use is dependent on the purpose. The membership function choice is the subjective aspect of fuzzy logic, it allows the desired values to be interpreted appropriately. Membership function that is used in this paper is trapezoidal function , with its L- and R- special cases.

Trapezoidal membership function – parametrized by a, b, c and d, where its generalized formula is given by equation (1) and also shown graphically in Figure .

$$\mu_{Trapezoidal}(x) = \max\left(\min\left(\left(\frac{x-a}{b-a}\right), \left(\frac{d-x}{d-c}\right), 1\right), 0\right) \quad (1)$$

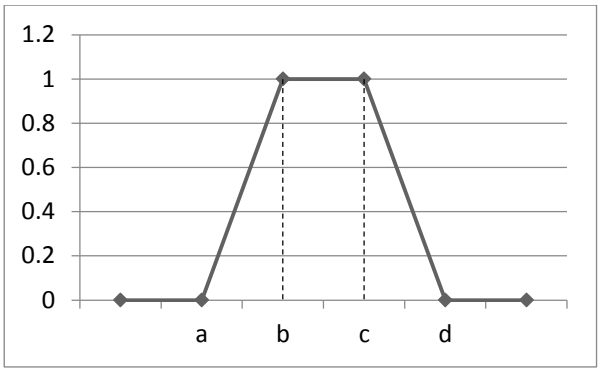


Figure 1 – trapezoidal membership function

L-trapezoidal membership function – special case of trapezoidal membership function, parametrized by a and b, where its generalized formula is given by equation (2) and shown graphically in Figure .

$$\mu_{L-trapezoidal}(x) = \max(\min\left(\left(\frac{b-x}{b-a}\right), 1\right), 0) \qquad (2)$$

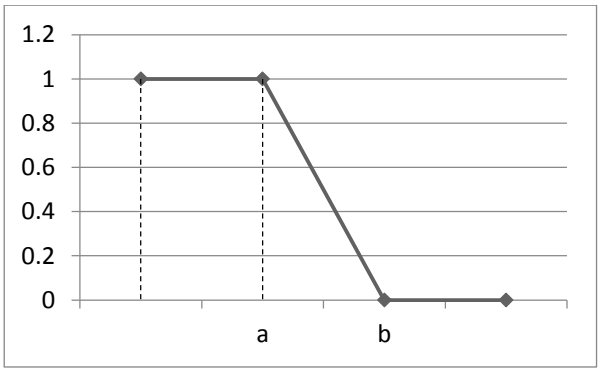


Figure 2 - L-trapezoidal membership function

R-trapezoidal membership function – special case of trapezoidal membership function, parametrized by a and b, where its generalized formula is given by equation (3) and shown graphically in Figure .

$$\mu_{R-trapezoidal}(x) = \max(\min\left(\left(\frac{x-a}{b-a}\right), 1\right), 0) \qquad (3)$$

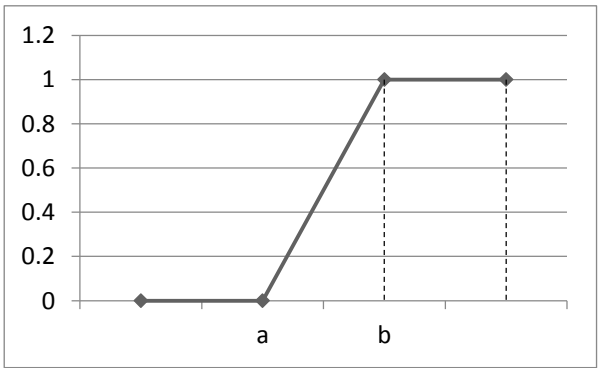


Figure 3 - R-trapezoidal membership function

B. *Graph API*

Facebook Graph API is the primary way to read and write to the Facebook social graph. The Graph API has multiple versions available, and in this research version v2.1 is used. By using this API, it is possible to do lots of different queries, e.g. fetching data about certain events, apps, groups, pages, users, and also user-specific data like user's timeline, friends, photos, etc. Full list of root nodes of the Graph API version v2.1 is available on https://developers.facebook.com/docs/graph-api/reference/v2.1.

In order to use the Graph API, one must provide a valid access token in each request, which is gained when someone connects with an app using Facebook Login, and this authentication flow is based on the OAuth 2.0 protocol. The received access token is so-called short-lived access token, and it expires after approximately 2 hours. However, this short-lived access token can be exchanged with long-lived access token, which expires after 60 days.

C. *Collecting the data used in this research*

We have used Facebook Graph API in order to download data for around 1.3 million Facebook pages that are analyzed in this research, and this downloading is done on daily basis by using obtained 300 access tokens. The data consists of basic information about each page (name, description, category, likes count, etc.), and edge for this data is simply **/page/{PageId}**, where {PageId} is a unique identifier for each page. Also, posts by each page are downloaded, where each post carries information how much users liked it, commented on it and shared it. The edge for this data is **/page/{PageId}/posts?fields=likes.limit(1).summary(true),comments.limit(1).summary(true),shares**

IV. PROPOSED METHOD

Determining the measure of influence of a Facebook page consists of three different sub-measures: 1) total number of fans of the page, 2) page activity, i.e. how often page posts a status, picture, video or link, and 3) total engagement of fans per page post. Total engagement per post is calculated using the number of likes, comments and shares of that post.

Number of fans

We differentiate 5 types of Facebook pages by total number of fans criteria: very small, small, medium, large, and very large. Each of these types has its corresponding fuzzy set. Membership functions for these fuzzy sets follow. Argument **n** represents the number of fans.

- Very small number of fans is described by L-trapezoidal membership function $\mu_{F,VerySmall}(n)$, whose parameters are $a = 100,\ b = 1000$
- Small number of fans is described by L-trapezoidal membership function $\mu_{F,Small}(n)$, whose parameters are $a = 1000,\ b = 10000$

- Medium number of fans is described by trapezoidal membership function $\mu_{F,Medium}(n)$, whose parameters are $a = 7500,\ b = 10000,\ c = 100000,\ d = 125000$
- Large number of fans is described by R-trapezoidal membership function $\mu_{F,Large}(n)$, whose parameters are $a = 100000,\ b = 500000$
- Very large number of fans is described by R-trapezoidal membership function $\mu_{F,VeryLarge}(n)$, whose parameters are $a = 1000000,\ b = 5000000$

Page activity

Page activity is defined as an average number of page posts per day in the last arbitrary N_D days. By this criterion, we separated pages into three categories: inactive, active, and very active. Each of these categories is described by its corresponding fuzzy set. Argument **p** represents an average number of page posts per day in the last N_D days.

- Inactive pages are described by L-trapezoidal membership function $\mu_{A,Inactive}(p)$, whose parameters are $a = 0.2,\ b = 0.5$
- Active pages are described by trapezoidal membership function $\mu_{A,Active}(p)$, whose parameters are $a = 0.5,\ b = 1,\ c = 4,\ d = 8$
- Very active pages are described by R-trapezoidal membership function $\mu_{A,VeryActive}(p)$, whose parameters are $a = 5,\ b = 15$

Here, we need to note that if **p** has value less than 1, it doesn't sound very reasonable in natural language, e.g. page has 0.5 posts per day, and of course it is impossible to post only half of the post. In order to understand this better, we can take reciprocal value of this number and interpret it as an average number of days that need to pass for page to publish a post, e.g. average 0.5 posts per days means that one needs to wait on average for 2 days to see new post of that page.

Total engagement

Total engagement basically represents the measure of popularity of a certain post, and also amount of interaction on that post. It is calculated by using total number of likes, comments and shares. However, comments represent higher amount of fan interaction then by simple like, while share being the highest level of interaction (share propagates the post even more through user's timeline). Because of these facts, we introduce weights k_1, k_2, k_3, where $k_1 \leq k_2 \leq k_3$ are chosen when calculating total engagement.

Final value is calculated by dividing the weighted sum of likes, comments and shares by total number of fans:

$$e = \frac{k_1 * likes + k_2 * comments + k_3 * shares}{fans}$$

This measure is more useful because it gives relative value of total engagement.

Facebook pages are divided into 3 groups by an average value of relative total engagement for posts in the last arbitrary N_D days: unpopular, popular, and very popular.

- Unpopular posts are described by L-trapezoidal membership function $\mu_{E,Unpopular}(e)$, whose parameters are $a = 0.05,\ b = 0.15$
- Popular posts are described by trapezoidal membership function $\mu_{E,Popular}(e)$, whose parameters are $a = 0.1,\ b = 0.2,\ c = 0.4,\ d = 0.6$
- Very popular posts are described by R-trapezoidal membership function $\mu_{E,VeryPopular}(e)$, whose parameters are $a = 0.3,\ b = 0.5$

It is also important to note that some users that are not fans of a certain page could like, comment or share that page's posts. However, it is not very likely, and using the proposed fuzzy approach indeed becomes very useful when modelling data with such imprecision.

Determining measure of influence

Using the proposed measures, which are described as fuzzy sets, it is possible to find Facebook pages by different criteria. To do this, one has to multiply values of membership functions for each criterion, as given in the following equation:

$$\mu = \mu_F(n) * \mu_A(p) * \mu_E(e)$$

Note that $\mu_F(n)$ component is a measure of influence *quantity* (it just represents number of fans, which solely isn't enough to determine influence of a page), while $\mu_A(p) * \mu_E(e)$ component is a measure of influence *quality* (it represents how audience actually responds to page activities). If certain page doesn't post often, but the engagement is high, it is probable that fans just had plenty of time to like these infrequent posts. On the other hand, if a page posts lot of posts and users don't respond much to it, then it is likely that users don't find that content to be relevant. However, pages with high frequency of posting, along with high engagement on the posts are definitely interesting to be considered as influential candidates.

For example, with this method, one can find active Facebook pages that have very large number of fans, and whose posts are popular among those fans. Calculation of value $\mu = \mu_{F,VeryLarge}(n) * \mu_{A,Active}(p) * \mu_{E,Popular}(e)$ for each page in a given database is done, and then these records are sorted in descending order, thus getting the most relevant records first. In addition to this, it is also possible to filter these Facebook pages by its Facebook category (company, organization, magazine, brand, product, movie, music, TV, sport, website, blog, and many more, along with subcategories).

V. RESULTS

In this chapter we show the results of applying the proposed method to the data on 1.3 million Facebook pages. For measuring page activity and posts engagement, all posts in the last $N_D = 28$ days are taken. Also, for calculating total engagement, values for constants k_1, k_2, k_3 are $k_1 = 1, k_2 = 2, k_3 = 4$ due to shares being highest level of user's interaction with post, while comments represent lower level of interaction then share, but higher then likes. Few examples with top 5 results follow.

Very active Facebook pages with very large number of fans and whose posts are popular are shown in Table 1.

Name	Category	Fans	Posts/ day	Total eng.
Angels for Animals Network	Community	217325	25	0.22
Boomchampionstt.com	Media/news/ publishing	266926	12.5	0.15
Hot FM Mackay	Media/news/ publishing	212848	6.2	0.24
Stereo Visión	Radio station	191496	6.2	0.22
98FM	Radio station	185748	6.2	0.17

Table 1 – very active Facebook pages with very large number of fans and whose posts are very popular

Active Facebook pages with small number of fans and whose posts are very popular are shown in Table 2.

Name	Category	Fans	Posts/ day	Total eng.
Raise Your State	Coach	3113	3.6	0.49
My Little People	Just for fun	3567	1.7	0.62
Peace For Paws	Community	4718	4.2	1.34
Quality for Life Philippines	Food/beverages	5349	2.3	1.06
Intersport Ylivieska	Outdoor gear/ sporting goods	5200	2.1	0.46

Table 2 - Active Facebook pages with small number of fans and whose posts are very popular

We can also find *active Facebook pages in "news/media website" category, with very large number of fans and whose posts are unpopular*. Results are shown in Table 3. These pages are typical examples of pages with a huge number of fans and low influence – their numerous fans simply don't care about the content they post.

Name	Category	Fans	Posts/ day	Total eng.
McDonald's	News/media website	55080143	1.3	0.01
One Direction Denmark	News/media website	3960021	2.3	0.01
Quiksilver	News/media website	3887983	2.1	0.005
Proud to be an American	News/media website	4617519	0.9	0.003
Country Music Nation	News/media website	3531147	3.6	0.007

Table 3 - active Facebook pages in "news/media website" category, with very large number of fans and whose posts are unpopular

Active Facebook pages in "journalist" category, with medium number of fans and whose posts are popular are shown in Table 4. These journalists can be characterized as influential, because even though they don't have large number of fans, they respond quite well to the content these journalists post.

Name	Category	Fans	Posts/ day	Total eng.
Eddo Bashir	Journalist	17162	0.8	0.53
Adrian Vrauko	Journalist	25813	2.3	0.11
Tobias Schlegl	Journalist	12887	0.6	0.13
Paulo Eduardo Martins	Journalist	83944	0.5	0.25
Marcelo Tas	Journalist	3360794	2.8	0.06

Table 4 - Active Facebook pages in "journalist" category, with medium number of fans and whose posts are popular

CONCLUSION

In this paper we proposed a method for determining measure of influence of Facebook pages. Fuzzy approach is used to describe and determine measure of influence of Facebook pages. Three characteristics of a Facebook page are measured: total number of fans of Facebook page, activity of Facebook page (frequency of posting) and post engagement of fans (weighted sum of number of likes, comments and shares) per post. For each of these measures fuzzy sets were created that model the data using natural language.

Combining these three measures, our method can be used to find influential or non-influential Facebook pages in certain category. These results can be useful for marketers when starting a campaign, because they would be able to find out which Facebook pages could be suitable and Facebook pages should be avoided for that campaign.

Further advancement of the proposed method could be adding the language and regions filter, so that one could find pages only for certain locale. Also, analyzing posts in order to determine their sentiment (positive, negative or neutral) could be a viable direction of further research.

REFERENCES

[1] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman, "Influence and passivity in social media," in *Machine learning and knowledge discovery in databases*, Springer, 2011, pp. 18–33.

[2] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi, "Measuring User Influence in Twitter: The Million Follower Fallacy.," *ICWSM*, vol. 10, pp. 10–17, 2010.

[3] "Graph API," *Facebook Developers*. [Online]. Available: https://developers.facebook.com/docs/graph-api. [Accessed: 27-Dec-2014].

[4] "Social Media Marketing, Statistics & Monitoring Tools," *Socialbakers.com*. [Online]. Available: http://www.socialbakers.com/. [Accessed: 28-Dec-2014].

[5] "Simply Measured | Easy Social Media Measurement & Analytics," *Simply Measured*. [Online]. Available: http://simplymeasured.com/. [Accessed: 28-Dec-2014].

[6] "Social Media Monitoring Tools & Sentiment Analysis Software," *Trackur*. [Online]. Available: http://www.trackur.com/. [Accessed: 28-Dec-2014].

[7] "Klout | Be Known For What You Love," *Klout*. [Online]. Available: https://klout.com/home. [Accessed: 28-Dec-2014].

[8] Đ. Obradović, Z. Konjović, E. Pap, and N. M. Ralević, "The maximal distance between imprecise point objects," *Fuzzy Sets and Systems*, vol. 170, no. 1, pp. 76–94, May 2011.

[9] D. Obradovic, Z. Konjovic, E. Pap, and I. J. Rudas, "Modeling and PostGIS implementation of the basic planar imprecise geometrical objects and relations," in *Intelligent Systems and Informatics (SISY), 2011 IEEE 9th International Symposium on Intelligent Systems and Informatics*, 2011, pp. 157–162.

[10] D. Obradovic, Z. Konjović, and M. Segedinac, "Extensible Software Simulation System for Imprecise Geospatial Process," presented at the ICIST, Kopaonik, 2011, pp. 1–6.

[11] D. Obradovic, Z. Konjović, E. Pap, and I. J. Rudas, "Linear Fuzzy Space Based Road Lane Model and Detection," *Knowledge-Based Systems*, Jan. 2012.

[12] D. Obradovic, Z. Konjovic, E. Pap, and M. Jocic, "Linear fuzzy space polygon based image segmentation and feature extraction," in *Intelligent Systems and Informatics (SISY), 2012 IEEE 10th Jubilee International Symposium on*, 2012, pp. 543 –548.

[13] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.

A Framework for Comparative Analysis of Data Mining Algorithms

Duško Mirković*, Ivan Luković**, Nikola Obrenović*, Đurđa Rogić*
*Schneider Electric DMS NS LLC, Novi Sad, Serbia
** Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
{dusko.mirkovic, nikola.obrenovic, djurdja.rogic}@schneider-electric-dms.com, ivan@uns.ac.rs

Abstract—Consumer load profiles play an important role in distribution network analysis. Determining typical consumers can be based on a data mining algorithm called clustering or cluster analysis. Given the multitude of clustering algorithms available today and the disparity of data sets, algorithm selection becomes a non trivial task. One approach to this task is to use multi-criteria decision making algorithms to rank data mining algorithms based on performance and other relevant metrics. In order to move focus from algorithm ranking to selection and tuning, one needs a framework that offers performance data manipulation as well as flexible and customizable ranking algorithms. This paper proposes one such framework that will support research of consumer clusterisation algorithms for power distribution networks.

I. INTRODUCTION

One of the main characteristics of electric energy is that it cannot be directly stored in relevant quantities from the standpoint of distribution systems. Electric energy is cheapest when it is produced in large quantities in facilities such as coal or nuclear power plant. However, these power plants have relatively high response time for changing the output power and starting or stopping may take several days. More responsive power plants are subject to limitations such as geographical location (e.g. hydroelectric power plants require water reservoir such as natural or artificial lake), available capacity (e.g. water level in reservoir), high marginal energy cost, environmental issues etc. Indirect methods for energy storage, such as pumped hydroelectric (PHES) or compressed air (CAES) energy storage, are also subject to certain limitations such as geographical location (e.g. PHES requires two reservoirs at different heights), capital cost, power rating, available capacity and environmental issues [1]. Exact consumer needs at any given time are, in many cases, poorly predictable. Some of the main causes are the multitude of ways electric power is used today as well as number of non-industrial consumers. Home generation units such as photovoltaic cells introduce even more variance. The balance between production and consumption is the reason behind load forecast - a process of estimating future consumer needs.

Typical distribution network consists of thousands of nodes and branches. Having a measurement of relevant physical quantities for each node and branch is not economically justifiable, due to the high price of smart metering devices. However, determining the state of the network is crucial step that is providing necessary input for all other calculations and analyses.

Individual consumer load is a stochastic variable. This makes it very hard to develop a model for each individual consumer. To overcome this uncertainty consumers are grouped based on similar demand and assigned a load profile that represents average consumer in that group. Data mining algorithms that can discover such groups are called clustering or cluster analysis algorithms. Load profiles created this way are used in power distribution system for the processes of state estimation and load forecast.

Nowadays there are many algorithms for data clustering and classification. One of the main issues in any research that is based on data mining algorithms is selecting the algorithm which will give the best results for a given data set. Rice in [2] presented a base for algorithm selection problem in general, based on approximation theory. Wolpert and Macready showed that all algorithms that search for an extremum of a cost function perform exactly the same, according to any performance measure, when averaged over all possible cost functions [3]. Dubes and Jain compared several clustering algorithms from the user's perspective and concluded that rational basis for comparing clustering methods is needed with links to well-understood mathematical and statistical methods [4].

Conclusions of Wolpert and Macready in [3] and Dubes and Jain in [4] imply that data set characteristics are tightly coupled with performance of particular algorithm and play an important role in algorithm selection. However, handling that data and ranking should not take much effort that would be better spent trying new algorithms or tuning existing ones. Therefore, a framework is needed to handle all those tasks and allow researchers to focus on experiment.

This paper proposes a framework that provides benchmarking and a comparative analysis of data mining algorithms with regard to data set characteristics as well as all relevant performance metrics. The framework is to provide flexible data model and an extensible process for performance indices collection. Collected data are used to rank algorithms by one or multiple Multi-Criteria Decision Making (MCDM) algorithms.

Such framework will serve as a workbench for further research in selecting the most appropriate algorithm for consumer clustering in power distribution systems.

Beside the Introduction and Conclusion, this paper consists of three sections. In section 2 we present related work. In Section 3 we analyze main requirements for the framework we present in this paper. The framework architecture is presented in section 4.

II. RELATED WORK

This section presents related work from three aspects: existing MCDM software, use of MCDM for data mining algorithm ranking (selection) and MCDM algorithms.

International Society on Multiple Criteria Decision Making[1] offers an extensive list of MCDM related software that we considered for our research. However, each of the solutions that we considered had some limitations that drove us to the need to develop a new framework that will support our and possibly many other researches in the field of data mining. Most of the solutions that we considered only supported one MCDM algorithm which can be limiting for research teams that want to try different algorithms to find the one that is most suitable for their actual research. Our framework aims to provide extensible model that allows virtually any MCDM algorithm to be plugged in and tested against collected data.

Another limitation that we encountered is that MCDM solutions require manual entry of all alternatives and their attributes. This may be necessary for project portfolios where none of the attributes can be gathered automatically. Our framework aims to provide means to describe a workflow. For each step of the workflow performance data is collected automatically. Also, each step of the workflow is modular so we can easily explore variations in performance by substituting only one step of the workflow. Solutions that we considered were either web (cloud) based or standalone tools. Our framework aims to provide easy collaboration mechanism such as peer to peer, without the need for central server host.

Microsoft offers a cloud based environment for data mining - Azure Machine Learning[2], which provides means for result evaluation. However, to the best of our knowledge it does not provide any MCDM algorithms that would help in selecting appropriate solution from several non-dominated solutions.

We believe that software that would fulfill most of the requirements for many disparate research projects would be hard to build and would take too much effort. On the other hand, our framework provides basic building blocks that can be used to build custom tailored solutions. It also provides referent implementation that can be either directly used or adapted for particular situation. This means that each research project has flexible starting point that does not require experienced software engineer to customize according to project specifics. Common components allow better knowledge sharing and exchange of experience.

One of the first attempts of using MCDM approach to solve the users' dilemma for selection of data mining algorithm was based on Data Envelopment Analysis (DEA) [5], a method in operations research and economics for measuring productive efficiency of decision making units. This approach is later extended with user profiles that would give more importance to some parameters in order to express user preference [6]. There were also attempts to use multiple algorithms simultaneously [7], [8]. This should provide better stability of the ranking and alleviate single algorithm weaknesses. Our framework is to provide abstract component that will represent both simple and hybrid MCDM algorithms.

Approach that was used in ESPIRIT METAL project was based on similarity to known data set performance [9], [10]. Data set characteristics and performance data were gathered and algorithms ranked using MCDM methods. This data was later used to estimate performance of algorithms for unknown data set. Our framework allows researchers to maintain custom attributes and use the collected data to perform such tasks. Another approach considered multiple human experts from different domains and modeled their preference with fuzzy sets that mapped qualitative expressions to weights that are used in selected MCDM [11], [12]. This approach is supported in our framework by means of custom MCDM modules that can implement any simple or hybrid approach.

The task of algorithm selection lies at intersection of many disciplines so it is surprising how little intersection has been in the relevant developments in different communities [13]. With each community developing its' own vocabulary it was harder to search for relevant papers and build on existing work. Keogh and Kasetty chose 57 of the most relevant papers at the time and re-tested them against 50 diverse data sets [14]. They showed that most of the benchmarks in the field of data mining algorithms were performed on very limited data set without explicit note and this can easily lead to false conclusions about performance. In some cases small variations in implementation of known algorithm gave better performance improvements than the newly proposed algorithm. For this reason it is particularly important to precisely state the characteristics of data sets that are used in benchmark and perform unbiased optimization of the algorithm.

MCDM algorithms are one way to formally define decision making process and thus minimize bias towards certain solutions as well as provide more information for someone that is looking at our conclusions. With this formal definition of our decision process, interested reader can compare our preferences with his and have better idea of how relevant our conclusions are compared to his specific problem. However, collecting of the relevant data and ranking alternatives by one or more MCDM algorithms requires nontrivial effort. This is why we decided to create a framework that will make this process easier and will help in knowledge sharing by providing more details and common nomenclature.

Our framework does not aim to provide exhaustive set of MCDM algorithms, but rather enable flexible interface that will allow virtually any algorithm to be plugged in. In order to derive a common algorithm interface, we studied several of the most widely known MCDM algorithms. Algorithms that were studied are: Weighted Sum Model (WSM) and similar but less used Weighted Product Model (WPM), Preference Ranking Organization Method for Enrichment of Evaluations (PROMETHEE) [15], Višekriterijumska Optimizacija i Kompromisno Rešenje (VIKOR) [16]–[18], Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) [19], [20], Logic Scoring of Preference (LSP) [21]–[24], Data Envelopment Analysis (DEA) [25]–[27], Adjusted Ratio of Ratios (ARR), and Analytic Hierarchy Process (AHP) [28].

[1] http://www.mcdmsociety.org/ available in December 2014.
[2] http://azure.microsoft.com/en-us/services/machine-learning/ available in December 2014.

III. REQUIREMENTS

This section presents main requirements that will drive the framework architecture and design. We first describe typical process for clustering consumers in power distribution system. We then abstract the steps where possible so that the requirements we identify may be attributed to general data mining task.

Typical process of clustering consumers in power distribution system consists of 5 major steps: input data cleansing, populating consumer model, dimensionality reduction, clustering and cluster evaluation. We outline these steps in Fig. 1. More details about each of the steps is given in the following paragraphs.

The first step is cleansing of the input data. Input data for consumer clustering is collection of measurements of active and reactive power at regular intervals (usually 15 minutes) for one or several years. It is possible to have missing data or peaks. Peaks are unusually large values that are caused by a network disturbance or a measuring equipment malfunction. Both missing values and peaks can have negative effect on clustering algorithm. There are several strategies to resolve such situations, such as interpolation, but they are out of scope of this paper.

Almost any data mining task based on real world data will require such step. Type of irregularities may differ but conceptually this step remains the same: processing step that requires certain amount of resources and changes the quality of the input data which can have non-trivial effect on further processing steps. Resources that it uses as well as data quality change may be measured and used to select the most appropriate algorithm for data cleansing for particular workflow.

The second step is populating the consumer model, which is explained in details in the remaining of the section. Two consumers behave similarly not only if they draw similar power from the network at one moment in time, but rather on the entire interval (e.g. one or multiple years). However, comparing all measurements of two consumers, for the given period of one or multiple years, would produce complex model with too many details that would make it hard to identify groups of similar consumers, due to phenomenon called *dimensionality curse* [31]. Vladimir Pestov in [29] discusses one aspect of this phenomenon: a point in high-dimensional space can have many "close" neighbors. This is much wider subject and there are many more relevant papers that deal with it, but it is out of scope of this paper.

In order to partially avoid the dimensionality curse, first level of abstraction is introduced, a consumer model. A consumer models consists of a set of curves where each curve represents consumer's consumption during 24 hours, for a particular season (e.g. Spring, Summer, Autumn and Winter) and a particular day type (e.g. working day, weekend or holiday). Each curve from the consumer model is calculated as an average from consumer's daily curves which correspond to the given season and the given day type. By using the consumer model to represent a consumer, we raise the level of abstraction and move away from the detailed measurements. Also, the aggregation of detailed measurements reduces the influence of missing or invalid data, that missed to be cleaned in the previous step.

In the previous step we reduced consumer representation to approximately few thousand dimensions

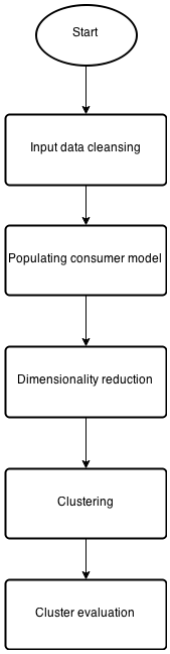


Figure 1 Typical process of clustering consumers in power distribution system

(192 measurements per day, 3-5 day types, 4 seasons) which is still an order of magnitude less than direct representation (192 measurements per day, 365 days per year). However, this is still enough to manifest the aforementioned dimensionality curse. This is why another step is taken to reduce a number of model dimensions.

The third step is dimensionality reduction, which applies one of many algorithms to transform input data into another problem space with fewer dimensions. One of such algorithms is Principal Component Analysis (PCA) [30]. This algorithm is used to represent cross-correlation matrix of the consumer type model by several orthogonal vectors - principal components. We can either choose fixed number of principal components or variable number of principal components depending on their cumulative influence (e.g. top n principal components that will amount for 90% of the variation in data set). The second and the third step perform major compression of source data by creating more and more abstract models. Just for quick comparison we can presume that we had measurements at least for one year at 15 minutes intervals which amounts to approximately 70.000 measurements per customer. After the third step we represent the same data by not more than 60 data points per customer. It is obvious that these models should be selected carefully in order to provide good input data for further steps. The same as for the first step we can measure the quality of derived model as well as resources that were used in processing. This will later influence our decision on what algorithm to use and how to choose parameters, if any, for the selected algorithm.

The fourth step is clustering of the data represented in model from previous step. Each consumer is one data point represented in n-dimensional data space where n is the number of attributes in input model (e.g. if we used PCA in previous step then n is the number of principal components). There are many clustering algorithms available nowadays, but all of them have one thing in

common. They must have a metric defined on a problem space. Examples of such metrics are Manhattan distance, Euclidean distance, Chebyshev distance etc. [31]. Each of these algorithms uses different amount of resources depending on both the quantity of input data as well as metric that is chosen. They also produce results of different quality.

The last, fifth step is determining quality of the results when there is no ideal clustering to compare to. There are several widely used internal and external evaluation measures such as Davies-Bouldin index [32] and Rand measure [33]. Usually more than one is used to get a better comparison between results for different algorithms.

At this point we have some measurements of how much resources we have used and what is the quality of the results (both intermediate and final). The next question is if it is the most appropriate solution for this environment. To address this issue we would have to try some other approach and compare the results. As Keogh and Kasetty showed in [14], experiment result is strongly tied to data set characteristics as well as algorithm implementation on the specific platform. Resources that were used, input data quality and result quality will be analyzed and compared to alternatives to achieve a solution closer to global optimum. This is where MCDM algorithms can help with formally defined criteria and procedure for selection of one out of possibly many non dominated solutions.

The framework is to provide the components that wrap the steps described above as well as the input and output data sets. These components enable access to data that is used in decision making step, such as elapsed time, memory used, data quality etc. They also provide signaling and control flow operations that will enable creating the experiment workflow such as the one described above. This includes, but is not limited to, start step, step completed, cancel step, step error, etc. Actual step implementation is not included in wrapper component. In other words, wrapper component should be able to wrap around existing implementation of certain step (e.g. data cleansing). This is also valid for data set wrapper components. This enables use of the framework with many different databases, such as Hadoop, Vertica, SQL Server, Oracle, and many different languages, such as Java, C#, R, etc.

The framework is to provide abstract model of the entire consumer clustering process in such way that will enable implementation of at least following MCDM algorithms: WSM, WPM, PROMETHEE, VIKOR, TOPSIS, LSP, DEA, ARR, ELECTRE, AHP. The framework should also allow implementation of a custom MCDM algorithm that can be based on the same model of the consumer clustering process.

It is not unusual that more than one research team will work on one task such as selecting consumer clustering algorithm. Even if they belong to single organization such as corporation or several different organizations such as universities, they will possibly work on different platforms, such as Windows or Linux. The framework is to provide interface for collaboration for teams that use different platforms. In order to avoid one centralized location that would require special maintenance and administration, the framework is to provide distributed operation. In other words, there is no one central server that all the clients will connect to but rather every client can connect to any other client and form a network for collaboration. The framework is to define means for discovery of the first peer and all other peers already connected to it.

Collaboration communication may contain sensitive research data. The framework is to enable encrypted communication over public channels to protects sensitive research data from eavesdropping or content change. This should be modular, allowing research team to use encryption scheme that they consider appropriate for required data confidentiality. This includes no encryption scheme for situations where no confidential data is exchanged over public networks.

The framework is to define modular and loosely coupled architecture that will allow any component to be customized in a plug-in manner. In other words, research team should be able to provide custom implementation of any module and be able to use it without the need to recompile the entire workbench.

IV. THE FRAMEWORK ARCHITECTURE

In this section we propose an architecture of the framework. Main components are outlined in Fig. 2 and their characteristics presented from two broad aspects with regard to requirements stated in the previous section.

The first aspect is a support for the individual work of a researcher. This includes process modeling, execution and analysis of the results. The second aspect is a support for collaboration of researchers. This includes verification of results from other researchers as well as using their results to improve decisions about current research. For example, one researcher performs experiments with one algorithm and another researcher performs experiments on some other algorithm but the same or very similar input data set. They can share and compare experiment results in order to determine which algorithm is more appropriate.

Three major components provide researcher with functionality for design, execution and analysis of experiments. Process designer component provides modeling of the process as well as individual process steps. Each process step can be either nested process or primitive component. Primitive components are those that cannot be further decomposed, such as data set or algorithm. Primitive components have basic attributes that are measured and recorded such as number of data rows, average computation time, or used memory. Process has aggregated attributes that are calculated by aggregating process steps attributes. This enables two researchers to work on different layers of abstraction but still be able to compare the results. Such approach will effectively create attribute hierarchy that will define aggregation rules. For example: if we consider process that executes three sequential steps then we can simply sum the individual step elapsed time to get aggregated elapsed time; however, if the process executes three parallel steps then aggregated elapsed time will be equal to the maximum elapsed time of the three steps.

Each process step defines the implementation plug-in that will connect the process model with the actual executing process during the execution phase. Basic operations include, but are not limited to the start step, stop step, pause step, report progress (on demand or via callback method), set parameter, get parameter and get attribute. Get attribute is the only mandatory operation.

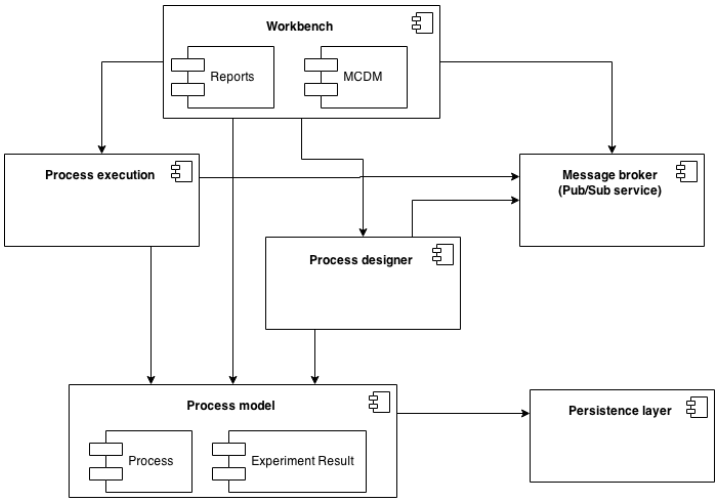


Figure 2 Proposed framework architecture

All others may or may not be supported by different implementations. For example: if we have offline trace processing plug-in it may only able to extract step attributes from execution trace but it will not be able to start or stop execution as it is already completed.

A result of the design phase is a data mining process model with all relevant attributes defined at each step. This model may be persisted and shared with other researchers. Model persistence is handled by a common component that provides basic operations such as store and retrieve, as well as model versioning. Model persistence includes execution results persistence. Each execution is uniquely identified and associated with process model and execution context. Execution context describes environment parameters such as platform, number of processors, available physical memory, as well as a user that executed the process.

The process execution component loads a process model from the persistent storage. Depending on the model it then offers supported operations to the user such as start processing or trace execution status. The common persistence component is used to record execution results as a part of the process model. Execution results contain values for step attributes defined in the process model. It is possible that not all attributes can be recorded for certain environments. In such case, researcher will be able to either create projection that does not contain those attributes with missing values or specify default values to use in the decision making.

Workbench is the third major component. It serves as a central point of individual work and binds all other components together. It is used to initiate design or execution sessions as well as provide modules for MCDM and reporting based on process model and experiment results. Workbench fills the process model based on messages it receives from other peers via message broker component that is described in more detail in the following paragraphs. The reports module provides many different views of the process model that help in analysis. The MCDM module provides plug-in port for implementation of various MCDM algorithms. A result of the MCDM algorithm is ranking or preference list of selected processes based on the experiment results. We rank processes characterized by algorithms used in each step so that we account for synergy between certain algorithms. The MCDM algorithms can also be used to rank algorithms used in a particular step of the process given that other steps are not changed in the selected experiments. These results are attached to process model together with context information and published to all subscribed peers.

The second aspect is a support for collaboration of researchers. The aforementioned workbench component uses the message broker component to provide pub/sub (publish/subscribe) functionality. Researchers create groups or teams on peer-to-peer principle. Each member of the group is able to invite new members. Each group member defines their level of interest that will dictate data that is exchanged. This prevents exchange of excessive data that could divert focus or increase pressure on communication channels.

Communication is based on the pub/sub principle. Each researcher subscribes to topic that are relevant for their research. Subscription requests are broadcasted to all peers. All peers also act as brokers, determining which publications are dispatched to which subscribers. Each peer dispatches only publications published on that peer. At any time subscriber can request special publication - integrity update. Integrity update publication transmits current state for the requested topic only to the requesting peer. It is used to initialize state of the subscriber after the subscription to certain topic as well as to reinitiate state after suspected communication failure. Integrity update request contains current state of the subscriber that the publisher can use to detect differences and send only data that is missing on the subscriber. Message broker component supports several protocols and communication channels in order to provide seamless collaboration even in the situations of complex network topology structures.

V. CONCLUSION

We started our research in the direction of finding the most appropriate consumer clustering algorithm. This search lead us to more elaborate problem of algorithm selection and MCDM problems. As Smith-Miles reported in [13] there was little intersection between relevant developments in different communities. This was mostly due to a different terminology in different problem domains and different communities.

We studied algorithm selection problems and MCDM use in data mining problems and identified the need for a

support in a form of a framework that will enable easier collection of characteristic performance data, as well as decision making by some of the widely used MCDM methods. Such framework is to support easier collaboration by common nomenclature and decision making process description. This allows researchers to formally state what are the important aspects of both data and algorithms in their problem domain and more efficiently communicate this information to fellow researchers. With our framework, we strive to give more confidence in published results and allow both result confirmation and further research based on those results.

We described our consumer clustering process and derived basic requirements for such framework that will help us in our further research. This requirements serve as the base on which main elements of the framework are defined. Focus of this paper is on the requirements. More detailed description of the framework is subject of the future work.

Our future work is directed to detailed design of the framework and reference implementation of the tool based on that framework. Further, we continue our research of the most appropriate algorithm for consumer clustering. It will serve as a proof of concept of our framework. The framework supports future development of many other data mining processes that will be based on large amount of data that is generated and collected by advanced distribution management systems, e.g. theft detection, outage prediction, predictive maintenance and many more.

ACKNOWLEDGMENT

The research presented in this paper was partially supported by Ministry of Education, Science and Technological Development of Republic of Serbia, Grant III-44010.

REFERENCES

[1] H. L. Ferreira, R. Garde, G. Fulli, W. Kling, and J. P. Lopes, "Characterisation of electrical energy storage technologies," *Energy*, vol. 53, pp. 288–298, May 2013.

[2] J. Rice, "The algorithm selection problem," *Adv. Comput.*, vol. 15, 1975.

[3] D. H. Wolpert and W. G. Macready, "No Free Lunch Theorems for Search," Santa Fe Institute, Feb. 1995.

[4] R. Dubes and A. K. Jain, "Clustering techniques: The user's dilemma," *Pattern Recognit.*, vol. 8, no. 4, pp. 247–260, Oct. 1976.

[5] G. Nakhaeizadeh and A. Schnabl, "Development of Multi-Criteria Metrics for Evaluation of Data Mining Algorithms.," *KDD*, 1997.

[6] G. Nakhaeizadeh and A. Schnabl, "Towards the Personalization of Algorithms Evaluation in Data Mining.," *KDD*, pp. 289–293, 1998.

[7] Y. Peng, G. Kou, G. Wang, and Y. Shi, "FAMCDM: A fusion approach of MCDM methods to rank multiclass classification algorithms," *Omega*, vol. 39, no. 6, pp. 677–689, Dec. 2011.

[8] G. Kou, Y. Peng, and G. Wang, "Evaluation of clustering algorithms for financial risk analysis using MCDM methods," *Inf. Sci. (Ny).*, vol. 275, pp. 1–12, Aug. 2014.

[9] J. K. Helmut Berrer, Iain Paterson, "Evaluation of Machine-Learning Algorithm Ranking Advisors," 2000.

[10] P. B. Brazdil, C. Soares, and J. P. da Costa, "Ranking Learning Algorithms: Using IBL and Meta-Learning on Accuracy and Time Results," *Mach. Learn.*, vol. 50, no. 3, pp. 251–277, Mar. 2003.

[11] A. Sanayei, S. Farid Mousavi, and A. Yazdankhah, "Group decision making process for supplier selection with VIKOR under fuzzy environment," *Expert Syst. Appl.*, vol. 37, no. 1, pp. 24–30, Jan. 2010.

[12] M. Noor-E-Alam, T. F. Lipi, M. Ahsan Akhtar Hasin, and A. M. M. S. Ullah, "Algorithms for fuzzy multi expert multi criteria decision making (ME-MCDM)," *Knowledge-Based Syst.*, vol. 24, no. 3, pp. 367–377, Apr. 2011.

[13] K. A. Smith-Miles, "Cross-disciplinary perspectives on meta-learning for algorithm selection," *ACM Comput. Surv.*, vol. 41, no. 1, pp. 1–25, Dec. 2008.

[14] E. Keogh and S. Kasetty, "On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration," *Data Min. Knowl. Discov.*, vol. 7, no. 4, pp. 349–371, Oct. 2003.

[15] J. P. Brans, P. Vincke, and B. Mareschal, "How to select and how to rank projects: The Promethee method," *Eur. J. Oper. Res.*, vol. 24, no. 2, pp. 228–238, Feb. 1986.

[16] S. Opricovic and G.-H. Tzeng, "Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS," *Eur. J. Oper. Res.*, vol. 156, no. 2, pp. 445–455, Jul. 2004.

[17] G.-H. Tzeng, C.-W. Lin, and S. Opricovic, "Multi-criteria analysis of alternative-fuel buses for public transportation," *Energy Policy*, vol. 33, no. 11, pp. 1373–1383, Jul. 2005.

[18] S. Opricovic and G.-H. Tzeng, "Multicriteria Planning of Post-Earthquake Sustainable Reconstruction," *Comput. Civ. Infrastruct. Eng.*, vol. 17, no. 3, pp. 211–220, May 2002.

[19] C.-L. Hwang, Y.-J. Lai, and T.-Y. Liu, "A new approach for multiple objective decision making," *Comput. Oper. Res.*, vol. 20, no. 8, pp. 889–899, Oct. 1993.

[20] Y. Lai, T. Liu, and C. Hwang, "Topsis for MODM," *Eur. J. Oper. Res.*, vol. 76, pp. 486–500, 1994.

[21] J. J. Dujmović and H. Nagashima, "LSP method and its use for evaluation of Java IDEs," *Int. J. Approx. Reason.*, vol. 41, no. 1, pp. 3–22, Jan. 2006.

[22] J. J. Dujmović and H. Bai, "Evaluation and comparison of search engines using the LSP method," *Comput. Sci. Inf. Syst.*, vol. 3, no. 2, pp. 31–56, 2006.

[23] J. J. Dujmović and H. L. Larsen, "Generalized conjunction/disjunction," *Int. J. Approx. Reason.*, vol. 46, no. 3, pp. 423–446, Dec. 2007.

[24] J. J. Dujmović, G. De Tré, and N. Weghe, "LSP suitability maps," *Soft Comput.*, vol. 14, no. 5, pp. 421–434, Jun. 2009.

[25] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *Eur. J. Oper. Res.*, vol. 2, no. 6, pp. 429–444, Nov. 1978.

[26] P. Andersen and N. C. Petersen, "A procedure for ranking efficient units in data envelopment analysis," *Manage. Sci.*, vol. 39, no. 10, pp. 1261–1264, Oct. 1993.

[27] A. Charnes, W. W. Cooper, B. Golany, L. Seiford, and J. Stutz, "Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions," *J. Econom.*, vol. 30, no. 1–2, pp. 91–107, Oct. 1985.

[28] T. L. Saaty, "How to make a decision: The analytic hierarchy process," *Eur. J. Oper. Res.*, vol. 48, no. 1, pp. 9–26, Sep. 1990.

[29] V. Pestov, "On the geometry of similarity search: Dimensionality curse and concentration of measure," *Inf. Process. Lett.*, vol. 73, no. 1–2, pp. 47–51, Jan. 2000.

[30] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemom. Intell. Lab. Syst.*, vol. 2, no. 1–3, pp. 37–52, Aug. 1987.

[31] P. N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Pearson Addison Wesley, 2006, p. 769.

[32] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979.

[33] W. M. Rand, "Objective Criteria for the Evaluation of Clustering Methods," *J. Am. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, 1971.

Graph Layout Algorithms and Libraries: Overview and Improvements

Renata Vaderna, Gordana Milosavljević, Igor Dejanović

Faulty of Technical Sciences, University of Novi Sad, Serbia
{vrenata, grist, igord}@uns.ac.rs

***Abstract*—This paper focuses on exploring the possibilities of applying graph drawing algorithms to lay out custom diagrams, with emphasis put on UML class diagrams. Implementing even the simplest of layout algorithms that would lead to acceptable results requires excessive knowledge of graph theory. For this reason, many developers have to rely on existing solutions. There are several open source Java libraries for graph drawing and analysis, but most of them come with certain problems and limitations making their integration with a separately developed graphical editor overly complex. To deal with those issues, we are developing another graph drawing and analysis library, called Grad.**

I. Introduction and Motivation

When developing modeling tools or expanding existing ones, a need to automatically lay out diagram elements in an aesthetically pleasing way might arise.

One such example is a lightweight UML class diagram editor implemented as a part of a larger tool called Kroki [1, 14]. Kroki enables users to create sketches of business applications using several embedded tools, thus enabling each participant to use their preferred way of development: mockup editor, command console, or the mentioned lightweight UML class diagram editor. On top of that, sketches can be imported from general purpose modeling tools. Furthermore, the class diagram editor should be capable of opening imported sketches and sketches created using other Kroki tools i.e. showing class diagrams which represent them. Changes made in Kroki's class diagram editor are immediately visible in the mockup editor and vice versa, which explains the need for automatically arranging newly created elements (packages, classes and links between them).

Implementing even the simplest of layout algorithms that would lead to acceptable results requires excessive knowledge of graph theory. Furthermore, simply deciding which class of graph drawing algorithms would be best suited for the given application can be challenging for those new to this area of mathematics. For these reasons, many developers would have to rely on existing solutions. There are many open source libraries which focus on graph drawing and provide implementation of certain layout algorithms. With a large number of excellent graph libraries for C/C++ and Python, it should be emphasized that only Java libraries will be considered in this paper, such as the popular JGraphX, JGraphT, JUNG and Prefuse.

Although providing a decent number of layout algorithms, all of these libraries primarily focus on graph visualization, thus making simply calling the desired algorithm and retrieving the results overly complex. In addition to this, it is very unlikely that all elements of a certain diagram would be of the same size, which would require usage of a layout algorithm which takes this into consideration. Some of the available solutions, however, do not. On top of that, many algorithms handle recursive links and multiple links between the same two elements quite poorly. However, such links frequently appear in class diagrams, so these problems cannot be ignored.

Having the mentioned limitations in mind, we are developing another open source graph drawing and analysis library, called Grad (GRaph Analysis and Drawing) [2]. Unlike the other ones, it puts a lot of emphasis on the ease of integration with other graphical editors and deals with the previously mentioned problems.

The rest of the paper is structured as follows. Section 2 explains the need for automatically applying layout algorithms within an existing graphical editor by shortly describing Kroki's lightweight UML editor and the requirements it had to fulfill. Section 3 gives a brief introduction to graph theory and graph drawing algorithms. Section 4 showcases some popular Java graph drawing and analysis libraries and points out some of the problems encountered when integrating them with separately developed editors. Section 5 presents Grad, a library being developed in order to address the most common integration issues. Finally, section 6 concludes the paper and outlines future work.

II. Kroki's Lightweight UML Editor

A lightweight UML class diagram editor was developed as a part of a tool named Kroki. Kroki is used for rapid prototyping and participatory development of enterprise applications based on mockups. It enables users to create sketches of applications using a mockup editor, command console, by importing models from general purpose modeling tools and by using the mentioned UML class diagram editor. Changes made in Kroki's class diagram editor should immediately be visible in the mockup editor and vice versa. Therefore, Kroki's UML class diagram editor had to fulfill some additional requirements.

Firstly, packages, classes and their attributes and methods and links established between them should have additional semantics as they need to represent certain elements of the sketches. This leads to the conclusion that simply being able to visualize the sketches as class diagrams isn't enough.

Secondly, it should be possible to open diagrams corresponding to sketches created using other Kroki tools than the UML editor with it. When doing so, there is no

data available regarding positions of the UML classes formed from certain elements of the sketch. Therefore, a layout algorithm must be automatically performed. Without that, users would have to lay out the diagrams manually. Since these diagrams can be quite large, placing all of the elements in the desired positions would drastically slow down the use of the Kroki tool.

III. Basic Graph Theory Concepts and Graph Drawing Algorithms

In the following section, a short introduction to graph drawing theory, as well as an overview of the most commonly used algorithms will be given.

A. Basic definitions

A graph (V, E) is an ordered pair consisting of a finite set V of vertices and a finite set E of edges, that is, pairs (u,v) of vertices. A path is a sequence of distinct vertices, $v_1, v_2,..., v_k$, with $k \geq 2$, together with the edges $(v_1, v_2),..., (v_{k-1}, v_k)$. A cycle is a sequence of distinct vertices $v_1, v_2,..., v_k$, with $k \geq 2$, together with the edges $(v_1, v_2),..., (v_{k-1}, v_k), (v_k, v_1)$ [3] .

If edges are unordered pairs of vertices, then the graph is *undirected*. On the other hand, if edges are ordered pairs of vertices, the graph is *directed*. A graph is said to be *connected* if there is a path from any vertex to any other vertex in the graph. Graphs which are not connected are referred to as *disconnected*. Graphs which contain at least one cycle are called *cyclic* graphs, while the ones that do not are known as *acyclic*. A graph is *simple* if it doesn't contain any edges that join a vertex to itself (loops) or more than one edge connecting the same two vertices (multiple edges). Graphs which permit multiple edges are called *multigraphs*.

A drawing Γ of a graph G maps each vertex v to a distinct point Γ(v) of the plane and each edge (u, v) to a simple open Jordan curve Γ(u, v) with endpoints Γ(u) and Γ(v) [3]. A drawing is *planar* if no two distinct edges intersect except, possibly, at common endpoints. Some algorithms for constructing drawings of graphs are only designed for special classes of graphs, like trees, (simple, undirected, connected acyclic graphs), planar graphs (graphs which can be drawn in a plane without edges crossing), or directed acyclic graphs, while the other ones even work for general graphs.

B. An overview of graph drawing algorithms

An overview of the most popular classes of graph drawing algorithms will be given in the next couple of paragraphs. More detailed descriptions of them can be found in [3].

Tree drawing is one of the best studied areas of graph drawing. That is not surprising since automatic generation of drawings of trees finds many practical applications. All trees are planar, which means that it is always possible to construct drawings of them with no edge crossings. There are several time-efficient tree drawing strategies which allow creation of aesthetically pleasing drawings.

A circular drawing of a graph is its visualization with the following characteristics:

- the graph is partitioned into clusters
- the nodes of each cluster are placed onto the circumference of an embedding circle
- each edge is drawn as a straight line

These algorithms have many application, especially in tools that manipulate networks.

A rectangular drawing of a plane graph is a drawing of it in which each vertex is drawn as a grid point on an integer grid and each edge is drawn as a sequence of alternate horizontal and vertical line segments along the grid. These algorithms find applications in circuit layouts, database and entity-relationship diagrams and floorplanning.

Force-directed algorithms are among the most important classes of graph drawing algorithms. They are very flexible and can be used to calculate layouts of all simple undirected graphs. They calculate the layout of the graph using only information contained within the structure of the graph itself. Graphs drawn with these algorithms tend to be aesthetically pleasing, exhibit symmetries, and tend to produce crossing-free layouts for planar graphs. There are many force-driven algorithms, with Tutte's 1963 barycentric method being considered to be the first one. The most popular ones include Kamada and Kawai [4] and Fruchterman-Reingold [5].

Hierarchical drawing algorithms can be used when dealing with directed graphs (or *digraphs*) which represent hierarchies. Examples of hierarchies or near-hierarchies are, among others, class diagrams and function call graphs from software engineering. The main idea behind hierarchical methods is to modify force-directed methods to take into account edge directions, and use them to draw digraphs.

C. Class diagrams as graphs

Class diagrams can easily be viewed as graphs with the elements representing vertices and the links representing edges. They can have multiple links between the same two elements. On top of that, these diagrams can contain recursive links (links connecting one element to itself). Therefore, they can be viewed as multigraphs that can contain loops. These graphs can be planar, but there is no guarantee that that will be the case. The same goes for connectivity. Similarly, some class diagrams contain cycles, others are acyclic. Generally, class diagrams are directed, but this cannot be seen as a rule since links can, and often are navigable, but not always.

Having all of this in mind, as well as the descriptions of different types of graph drawing algorithms, it can be concluded that force-directed and hierarchical algorithms are most suitable for usage in class diagram layouters. However, simply performing these algorithms might not be enough to form an aesthetically pleasing drawing of a class diagram. Additional steps might have to be performed in order to show loops and multiple edges correctly.

IV. Related Work

There are quite a few libraries for analyzing and drawing graphs for Java. In this section, some of the most popular ones will be briefly described, with the focus being on quantity and quality of the implemented graph layout algorithms. Furthermore, a few examples of how these algorithms can be used by some other projects will be given, accompanied by a short discussion regarding the complexity of such calls.

A. A preview of the most widely used free graph libraries for Java

The most widely known and used free graph libraries include JGraphT [6] and JGraph [7], JGraphX [7], Prefuse [8] and JUNG (Java Universal Network/Graph Framework) [9]. It can also be noted that there are some commercial solutions, such as yFiles from yWorks [10], but, since using them in most projects is not a likely possibility, they will not be considered in this paper. Furthermore, neither will tools that only generate images, such as GraphViz [11], since their layout algorithms cannot be integrated with already existing graphical editors.

All of these libraries focus on enabling users to model and analyze and/or visualize data that can be represented as a graph or a network. Apart from JGraphT, all projects put heavy emphasis on visualization, some even allowing users to interact with the created graphs. In the following passages a short preview of some of the most popular Java graph libraries will be given, followed by a discussion regarding how convenient or inconvenient it would be to integrate their layout algorithms with already existing tools for visualizing the given data.

JGraphT is a free Java graph library that provides mathematical graph-theory objects and algorithms [6]. It enables simple graph creation and offers implementations of a wide range of graph analysis algorithms, such as Dijkstra's shortest path, but does not provide any layout algorithms. In fact, it relies on **JGraph** for visualization. A notable problem which users of this particular combination of libraries face is that JGraphT only supports usage of an older version of JGraph. JGraph was significantly enhanced and rewritten from scratch in version 6, with even the name being changed to JGraphX [7]. However, JGraphT wasn't updated, still using the old version of the previously mentioned visualization library.

JGraphX is a Java Swing graph visualization library. It enables integration of interactive diagrams into larger Swing applications [7]. It is possible to customize certain properties of the graphs such as design of the vertices, labels of the edges, etc. Most importantly, it also provides a few layout algorithms meant to assist users in setting out their graph. Most notably, JGraphX implements one rather effective force-directed algorithm, a simulated annealing layout based on [12]. Moreover, it also provides implementations of a few different tree layouts.

Prefuse is another library set of tools for creating interactive data visualizations [8]. Its distinguishing feature is the ability to read data and create graphs directly from XML files and relational databases with only a line or two of code. When it comes to layout algorithms, Prefuse, like JGraphX offers a number of tree layouts, but also two force-directed ones, including the mentioned Fruchterman-Reingold.

Java Universal Network/Graph Framework, also known as JUNG, is a library that offers both the possibility of analyzing and visualizing graphs [9]. The current distribution of JUNG includes implementations of a number of algorithms from graph theory, data mining and social network analysis, but also provides a visualization framework. JUNG framework, while not containing the largest number of implementations of different layout algorithms out of the other mentioned alternatives, does implement more force-directed ones. In fact, JUNG provides an implementation of the previously mentioned Fruchterman-Reingold algorithm and a slightly modified version of it, as well as Kamada-Kawai. However, it is quite complex to set custom sizes of the JUNG graph vertices.

B. Integration with existing graphical editors

The problem which will be analyzed in this section is how to use layout algorithms provided by the mentioned libraries within an already existing graphical editor. To be more precise, within an existing class diagram editor, where sizes of the vertices play a significant role. With all of the graph drawing libraries putting strong emphasis on visualization, simply calling a layout algorithm and retrieving the results i.e. positions of the vertices and, if available, information about locations and shapes of the edges, can be quite complex.

Typically, in order to call a layout algorithm, it is necessary to provide an instance of the graph class, meaning that the application's data model has to be transformed into the suitable format. In addition to that, visualization components may have to be initialized, even though they won't be used. More importantly, implementations of layout algorithms and/or graph, vertex and edge classes have to be analyzed in search of a way of retrieving information about the vertices and edges following the execution of layout algorithms. Out of the mentioned libraries, JGraphX and JUNG provide the largest number and the most complex layout algorithms. For this reason, examples will cover integration with their algorithms.

It is worth mentioning that integration with Prefuse is even more complex. Prefuse enables simple creation of graphs directly from XML files and relational databases. While it is easy to see why these features could be put to good use in many projects, it is dynamical creation of graphs which is of importance in this particular case. That, however, is accomplished much harder. JGraphT - JGraph combination also won't be used in the examples, since it is now obsolete, like it was explained in the previous section.

Every diagram of Kroki's UML class editor contains a list of elements and links between them. Let's assume that prior to calling the layout algorithms, elements were already loaded into a list called *diagramElements*, while the links were all inserted into a list simply called *links*. An example of calling a JGraphX layout algorithm and retrieving positions of the vertices is shown in code listing 4.1.

Firstly, it can be noticed that creating graphs using already existing elements is a bit inconvenient as JGraphX graphs aren't parametrized and thus cannot contain vertices of any given class. Secondly, one must be quite familiar with how JGraphX works in order to get positions of the vertices once layout algorithms have finished calculating them.

Accomplishing the same using the JUNG framework is much simpler, which can be seen by analyzing code listing 4.2. However, it is necessary to initialize the visualization component in order to trigger execution of layout algorithms. Also, retrieving positions from layout algorithm after it was performed, while easy to do, is not well documented and can prove to be quite hard to discover.

```
mxGraph graph = new mxGraph();
graph.getModel().beginUpdate();
Object parent = graph.getDefaultParent();
Map<GraphElement, Object> elementsJGraphXVerticesMap =
        new HashMap<GraphElement, Object>();
try
{
    for (GraphElement element : diagramElements){
        Object jgraphxVertex = graph.insertVertex(parent, null,
                element, 0, 0, element.getSize().getWidth(),
                element.getSize().getHeight());
        elementsJGraphXVerticesMap.put(element, jgraphxVertex);
    }
    for (Link link : links){
        Object v1 = elementsJGraphXVerticesMap.get(link.getOrigin());
        Object v2 = elementsJGraphXVerticesMap.get(link.getDestination());
        graph.insertEdge(parent, null, null, v1, v2);
    }

}
finally{
    graph.getModel().endUpdate();
}
mxOrganicLayout jgraphxOrganic = new mxOrganicLayout(graph);
jgraphxOrganic.execute(parent);
for (Object vertex : elementsJGraphXVerticesMap.values()){
    mxIGraphModel model = graph.getModel();
    mxGeometry geometry = model.getGeometry(vertex);
    //finally, we can get the x and y coordiates
    System.out.println(geometry.getX() + ", " +  geometry.getY());
}
```

Code listing 4.1 Calling a JGraphX layout algorithm and retrieving the results

```
UndirectedSparseGraph<GraphElement, Link> graph =
        new UndirectedSparseGraph<GraphElement, Link>();

for (GraphElement element : diagramElements)
    graph.addVertex(element);

for (Link link : links)
    graph.addEdge(link, link.getOrigin(), link.getDestination());

FRLayout<GraphElement, Link> layouter =
        new FRLayout<GraphElement, Link>(graph);

//triggers layouting
new DefaultVisualizationModel<GraphElement, Link>(layouter);

for (GraphElement element : diagramElements){
    Point2D p = layouter.transform(element);
    System.out.println(p);
}
```

Code listing 4.2 Calling a JUNG layout algorithm and retrieving the results

Another limitation of the JUNG framework, which was already briefly mentioned, is the complexity of setting custom sizes of the vertices. A solution to this problem proposed in [13] includes the use of aspects and requires considerable knowledge of the framework. On the other hand, simply using the default sizes of the graph vertices when performing layout algorithms can ultimately lead to their overlapping. Elements of class diagram, of course, fall into this category, which limits direct applicability of the JUNG framework.

Furthermore, it must be stressed that, as discussed in the third section, class diagrams are multigraphs that can contain loops. If that is indeed the case, in addition to calling the algorithms of the chosen library, users would have to handle loops and set positions of the overlapping edges (which happens when the graph has several edges between the same two vertices) themselves. In addition to the already mentioned problems, many algorithms do not perform particularly well if disconnected graphs are passed to them. There is often too much free space between the disjoint parts of such graphs.

V. GRAPH ANALYSIS AND DRAWING LIBRARY (GRAD)

The main motivation behind the project was to implement a variety of graph analysis and drawing algorithms and enable very simple integration with already existing graphical editors, which includes the possibility of calling the layout algorithms and retrieving the results very easily, while also being able to specify certain properties of the vertices, such as their sizes. It is worth mentioning that the current version of Grad also offers implementations of several graph analysis algorithms, such as planarity testing and graph traversal. However, they are not the main focus of this paper and will not be described in more detail. All examples of diagrams that will be shown in this section were created using Kroki's lightweight UML editor.

A. Layouting implementation

At the moment, there are five different layout algorithms available: three force-directed ones, one circular and the so-called box layout, which places elements in a table-like structure. Special attention was given to graphs with loops and multiple edges, enabling any custom class diagram to be arranged in an aesthetically satisfying way, with no edges overlapping. The problem that was mentioned in the previous section, regarding disconnected graphs was also addressed. Parts of these graphs are arranged separately and positioned in such way that they are neither too far apart from each other, nor too close. In the following passages, examples of class diagrams arranged using different layout algorithms will be shown.

The three currently provided **force-directed algorithms** are Kamada-Kawai, Fruchterman-Reingold and the basic spring algorithm. After performing these algorithms, additional steps are taken to make sure that no vertices are overlapping. If distances between any two vertices are smaller than the specified limit, their positions are adjusted. Users can set the limit themselves before calling layout algorithms. If they do not choose to do so, a predefined value is used. An example of an arranged class diagram with recursive links and two links between classes “Panel6” and “Panel2” using Kamada-Kawai algorithm is shown in Figure 5.1.

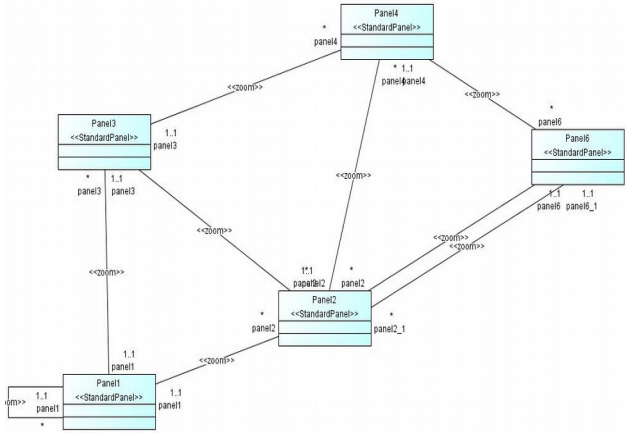


Figure 5.1 An example of a class diagram arranged using Kamada-Kawai algorithm

Like it was mentioned in section 3, force-directed algorithms tend to produce crossing-free layout for planar graphs. Looking at Figure 5.1, it can be noticed that this indeed is the case here. Also, the two links between the

same two classes don't overlap, and the recursive link connecting "Panel1" with itself is not hidden beneath the class.

Circular layout places vertices onto the circumference of an embedding circle. If the graph is biconnected (a graph which remains connected if any vertex is deleted), additional preprocessing is performed in order to minimize the number of crossings. The preprocessing involves calculation of the best possible order of vertices. An example of a class diagram arranged using the circular graph drawing algorithm is shown in Figure 5.2. Implementation of an algorithm for drawing non-biconnected graphs on multiple embedding circles is planned for future releases of Grad.

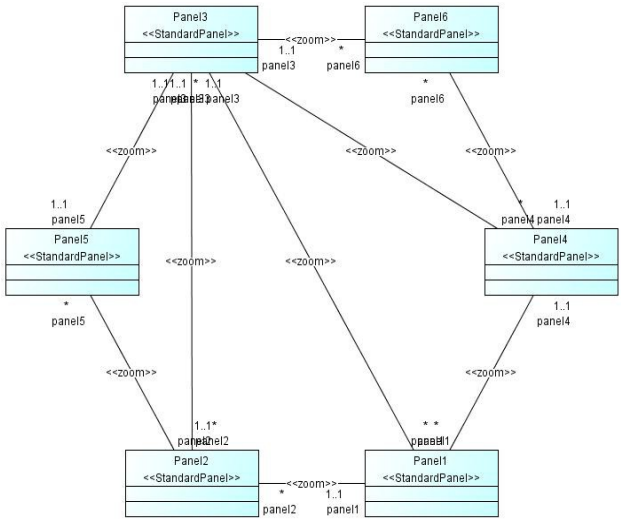


Figure 5.2 An example of a class diagram arranged using the circular graph drawing algorithm

Box layout places vertices in a table-like structure. The basic idea is to position a predefined number of vertices in one row, before continuing to the next one. Sizes of the vertices are taken into account when calculating heights of the rows and widths of the columns in order to prevent the vertices from overlapping. The number of vertices in a row can be adjusted by the user before executing the algorithm. If a class diagram is organized in such way that it contains a large number of packages on the first level, this layout is by far the best choice. An example of such usage is shown in Figure 5.3.

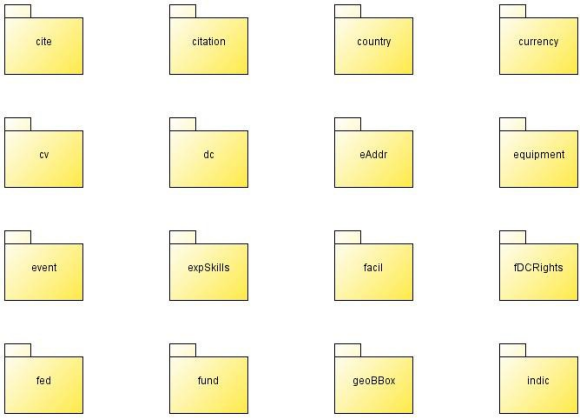


Figure 5.3 Using box layout to organize a diagram containing only packages

B. Integration with existing graphical editors

Grad can easily be used in combination with already existing graphical editors. In fact, the ease of integration was one of the project's main requirements.

The central class, which represents a graph, is parametrized, which means that it is safe to use just about any two classes as types of vertices and edges. The only requirement that must be fulfilled is that these classes have to implement appropriate interfaces (called *Vertex* and *Edge*), making it possible to easily specify properties of the vertices and edges (e.g. sizes of the vertices) which will later be used by the layout algorithms.

When calling a layout algorithm, one only needs to pass lists of vertices and edges, and not the already created graph since it will automatically be created later. Moreover, an object containing maps of vertices and edges and their positions is returned. Therefore, all information needed to position the editor's elements is obtained with no additional effort. Like it was already mentioned, Grad puts emphasis on properly handling loops and multiple edges. For this reason, edges can contain multiple segments whose endpoints are returned. A class diagram encapsulating previously explained is shown in Figure 5.4.

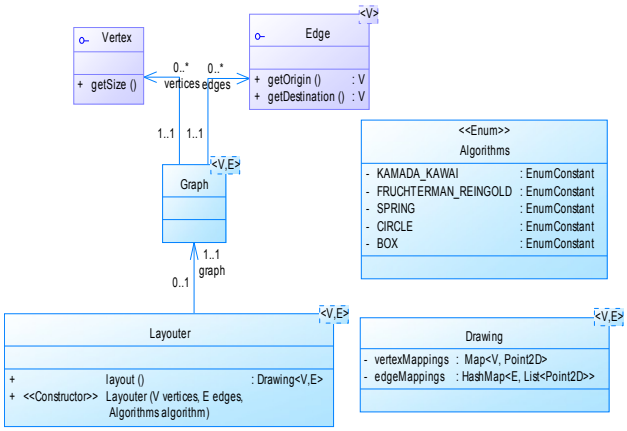


Figure 5.4 A part of Grad's model showing classes and interfaces needed for integration with other tools

It should be stressed that there is no need to dig deeper into Grad's implementation, to instantiate a visualization component or anything along those lines in order to call a layout algorithm and retrieve the results. To demonstrate that it is truly easy to do so, an example of calling the Kamada-Kawai graph drawing algorithm is shown in code listing 5.1. It is assumed that we have the same lists of elements and links like in chapter 4 at our disposal.

```
Layouter<GraphElement, Link> layouter =
        new Layouter<>(diagramElements,
                links, Algorithms.KAMADA_KAWAI);
Drawing<GraphElement, Link> drawing = layouter.layout();
Map<GraphElement, Point2D> elementPositions =
        drawing.getVertexMappings();
Map<Link, List<Point2D>> linkPosition =
        drawing.getEdgeMappings();
```

Code listing 5.1 Calling a Grad layout algorithm and retrieving the results

It can also be noted that it even isn't necessary to instantiate a class implementing the desired algorithm, as seen in code listing 5.1. The users only need to select an enumeration value which corresponds to their algorithm

of choice, while everything else will later be handled by the central class responsible for executing the layout algorithms, called *Layouter*.

VI. CONCLUSION

This paper presented an overview of different graph drawing algorithms and explored the possibility of integrating layout algorithms of some of the most popular Java graph drawing and analysis libraries with a separately developed graphical editor. In other words, in cases when their visualization isn't needed. Certain problems regarding ease of such integration, as well as some issues characteristic to arranging class diagrams were pointed out. They include the need to get quite familiar with the libraries before being able to call the layout algorithms they provide and retrieve the calculated positions of the vertices, as well as problems which might occur when the graph contains loops and multiple edges. These issues were addressed in a new graph drawing and analysis library called Grad.

Grad offers implementations of a variety of graph analysis and drawing algorithms, while focusing on the ease of integration with already existing graphical editors. Plans for future improvements of Grad include implementations of:

- an algorithm for drawing non-biconnected graphs on multiple embedding circles
- several tree drawing and hierarchical algorithms
- labeling algorithms which address automatic placement of text symbol labels

REFERENCES

[1] G. Milosavljevic, M. Filipovic, V. Marsenic, D. Pejakovic, I. Dejanovic, "Kroki: A mockup-based tool for participatory development of business applications", *IEEE 12th Conference on Intelligent Software Methodologies, Tools and Techniques*, pp. 235-242, 2013

[2] Graph Analysis and Drawing library (Grad), https://github.com/renatav/GraphDrawing, online, accessed January 11, 2015.

[3] Roberto Tamassia, *Handbook of Graph Drawing and Visualization*, Chapman & Hall/CRC, 2007.

[4] T. Kamada andS. Kawai, "An algorithm for drawing general undirected graphs", in *Information Processing Letters,* vol. 31, pp. 7-15, April 1989.

[5] T. Fruchterman and E. Reingold, "Graph drawing by force-directed placement" in *Software Practice and Experience,* vol. 21, pp. 1129 – 1164, November 1991.

[6] JGraphT, http://jgrapht.org, online, accessed January 11, 2015.

[7] JGraphX, https://github.com/jgraph/jgraphx, online, accessed January 11, 2015.

[8] Prefuse, http://prefuse.org, online, accessed January 11, 2015.

[9] JUNG Framework, http://jung.sourceforge.net, online, accessed January 11, 2015.

[10] yFiles, www.yworks.com/en/products/yfiles/, online, accessed January 11, 2015.

[11] GraphViz, http://www.graphviz.org, online, accessed January 11, 2015.

[12] Ron Davidson, David Harel, "Drawing Graphs Nicely Using Simulated Annealing", in *ACM Transaction on Graphics,* vol. 15, pp. 301-331, October 1996.

[13] Setting custom sizes of the JUNG graph vertices, http://sourceforge.net/p/jung/discussion/252062/thread/0b98adc5, online, accessed January 22, 2015.

[14] Kroki source, https://github.com/KROKIteam/KROKI-mockup-tool, online, accessed January 22, 2015.

Kroki Administration Subsystem Based on RBAC Standard and Aspects

Sebastijan Kaplar, Milorad Filipović, Gordana Milosavljević, Goran Sladić

Faculty of Technical Sciences, University of Novi Sad, Serbia
{kaplar, mfili, grist, sladicg}@uns.ac.rs

***Abstract*— This paper presents administration subsystem that was developed to enable dynamic customization of enterprise applications specified by our Kroki tool. Kroki is a tool for participative development of enterprise applications based on executable mockups. The administration subsystem is based on standard RBAC model for access control. It enables users to perform previously authorized tasks by dynamically adjusting their actions. Available actions are determined based on operations, relationships and constraints, according to RBAC. Every user role can access only specific parts of the application that its role is entitled to, with the use of menu adjusted especially for its needs. Dynamic adjustments of available actions are implemented by our runtime engines based on aspect-oriented approach.**

I. INTRODUCTION

Enterprise applications usually have a large number of users with different roles and responsibilities. For example, a worker in a warehouse can access only information about his warehouse. His superior can access information about all warehouses in the application. Financial director has the rights to access all available subsystems.

Enterprise applications should be dynamically adjusted to support specific needs of every user. In order to achieve this, we need an administration subsystem that allows specification of a working environment for each user (menu structure, user rights, etc.). Also, we need the architecture of the enterprise applications that can adapt to the mentioned specification on the fly.

This paper presents the administration subsystem that was being developed as a part of our Kroki tool [5, 7, 8] and generic engines that allow run-time enterprise application adaptation. Kroki (*croquis* – sketch) is an open-source tool for participatory development of enterprise applications based on executable mockups (see Section 3 for details).

The administration subsystem is based on RBAC (Role Based Access Control) standard for access control. RBAC introduces user roles as an additional layer of indirection between users and permissions. User roles can be created, modified, or deleted based on the enterprise system requirements, without the need to individually manage privileges for every user. RBAC-based systems enable users to perform previously authorized tasks, by dynamically adjusting their actions. Actions are determined based on operations, relationships and constraints [1].

The paper is organized as follows. Section 2 reviews the related work. Section 3 gives a short overview of our Kroki tool and describes the administration subsystem implementation. Section 4 presents an example of dynamic adjustments. Section 5 concludes the paper.

II. RELATED WORK

The paper [1] is one of the early papers that present a family of RBAC models. These models are provided as a common frame of reference for other research and development in the area, and they are used as a starting point for our work.

The paper [2] presents modeling and metamodeling of access control policies. The described process spans three meta-levels. At level M2, the policy metamodel is defined. Using the policy metamodel, different policy models can be applied at level M1, such as RBAC. PolicyDSL is used at level M0 for specifying actual access control policies in a particular system, and policy model is used for parameterizing the syntax of PolicyDSL.

In [3] the work is focused only on specifying the static structure of RBAC, and utilizes standardized modeling language (UML) and also integrates the policy specification activity with UML design modeling activities. Also, the task of capturing RBAC policies in reusable patterns is described.

In [4] SecureUML is introduced. SecureUML extends RBAC in order to add constraints on system states that are associated with a UML model. Rules are allowed to be restricted with OCL (Object Constraint Language) conditions. Also, action hierarchies allow forming of higher-level actions.

III. THE KROKI ADMINISTRATION SUBSYSTEM

The administration subsystem is developed as a part of our Kroki tool (Figure 1). Kroki is a tool designed for development of business applications based on executable mockups. A mockup is a sketch of an application UI (User Interface).

Kroki uses mockups as a basis for automatic execution and code generation for the enterprise applications. An advantage of mockup tools is in their ease of use, which allows end users to participate actively in the development process using simple and intuitive notation [9]. Besides mockups, Kroki also supports "classical" way of application modeling, using its lightweight UML editor [1].

Mockups execution is performed by two aspect-oriented (AOP) engines for the web and desktop applications. A basis for execution is application repository that contains configuration files for the current Kroki project. Configuration files are generated from the current model using Kroki generators (Figure 2). Although the engines could take this data directly from the

Kroki model, we choose XML files as an intermediate step in order to provide independent functioning of the specified applications, when deployed.

The administration subsystem is getting information about specified application elements (forms, panels, reports etc.) from the application repository. These elements are observed as resources in access control model.

Configuration files created by the administration subsystem are also delivered to the application repository. They are loaded during the engines start up and used as a source for dynamic adjustment of the application based on user rights.

Dynamic access to the user rights and their activation in appropriate moments is achieved with the use of AOP. Kroki engines use AOP techniques in order to enable easier integration of cross-cutting concerns imposed by its tools. Also, AOP enables easier integration of the engines with hand-written code which is necessary for implementation of complex business transactions and reports.

A tool for creating personalized menus is also a part of the administration subsystem. The tool helps in the process of creating custom menus, defined for every user role. Custom menu allows personalization of users working environment.

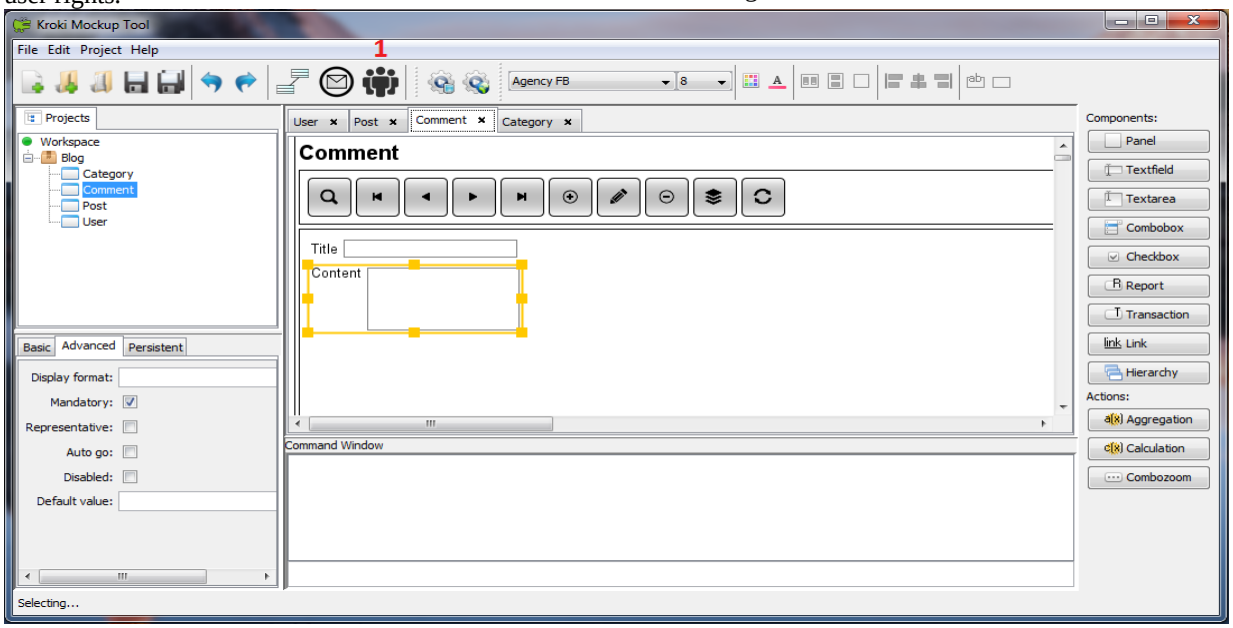


Figure 1. Kroki mockup tool. Icon for activation of the administration subsystem is marked with 1.

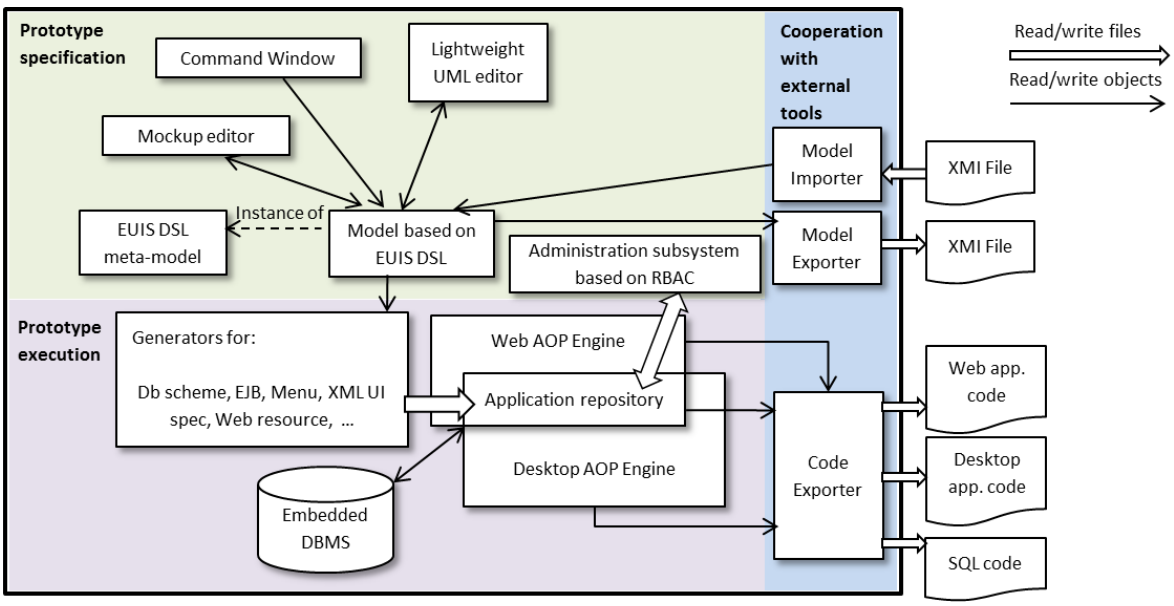


Figure 2. Kroki architecture

A. *RBAC*

The administration subsystem is based on RBAC, which is one of the most common access control models. RBAC is largely spread in business systems with a large number of users. The reason behind it is in the introduction of user roles. User role is an additional layer of indirection between users and permissions allowing the grouping of users and privileges in logical units at the higher level of abstraction. This enables simplified

specification of the user rights. RBAC model [3] is shown in Figure 3.

Basic entities defined with RBAC model shown in Figure 3 are sets of users, roles, objects, operations, and permissions. A role is a job function performed by the user. Object is an entity that contains or receives information. Permission is an approval to perform an operation (for example add, modify, remove) on the object. Permission is defined abstractly and implemented as a pair of permissions (objects, operations).

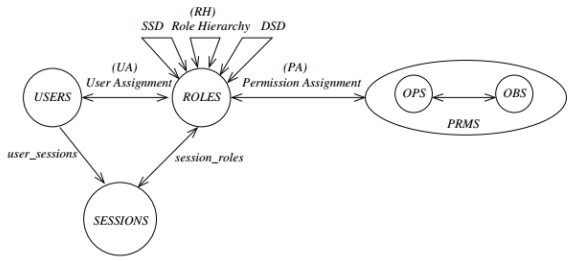


Figure 3. RBAC Model

The permission for execution of a certain operation is defined as a basic concept in RBAC. Associating permissions with roles simplifies their management. Users do not have direct permissions; they are obtained through roles.

Figure 4 shows the administration subsystem's UML model. In Figure 4, classes (entities) and their relationships in the administration subsystem are shown. One of the features of the model are resources, represented by the class *Resource*. Resources represent application forms sketched using Kroki's mockup editor. Storing data into resources is performed automatically through forms previously sketched in Kroki. Figure 4 shows that for each resource multiple permissions can be defined (*Permission* class) while each permission can have only one resource that results in one-to-many relationship.

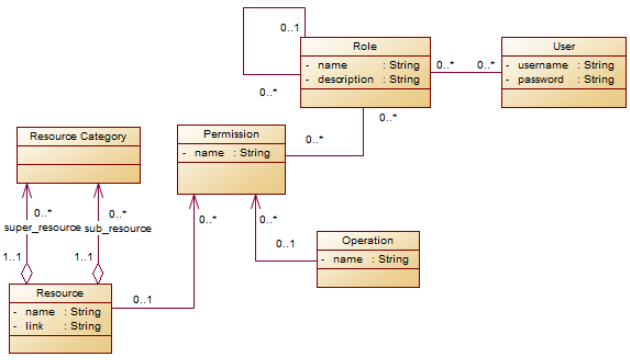


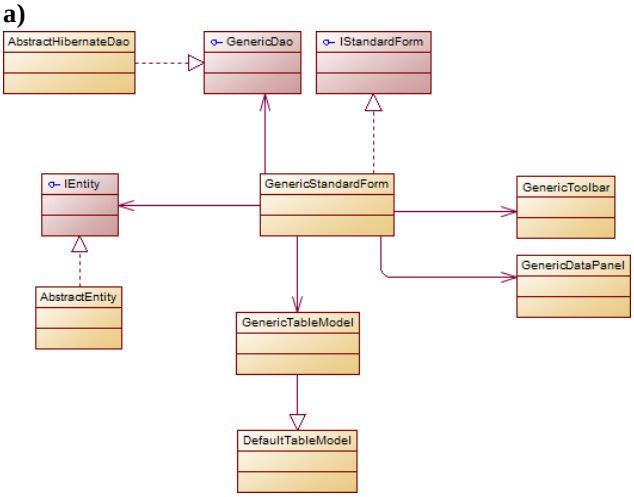
Figure 4. Administration subsystem model

Also, there is a categorization of the resources. Permission has one resource and one allowed operation (class *Operation*) on that particular resource, for example: add, modify, delete. For one resource, many permissions with different operations may exist and when a new operation is introduced it can add new permissions with the newly introduced operation on the existing resource. Roles (class *Role*) associate permissions with users. Figure 4 shows that one role can have more permissions, and one permission can have multiple roles that result in many-to-many relationship. Identical is the relationship between roles and users (class *User*).

B. *Implementation*

The administration subsystem is implemented as a three-tiered desktop application. A presentation part is implemented in Java using the Swing GUI library. The middle tier, responsible for business logic, uses Hibernate library for persistence.

a)



b)

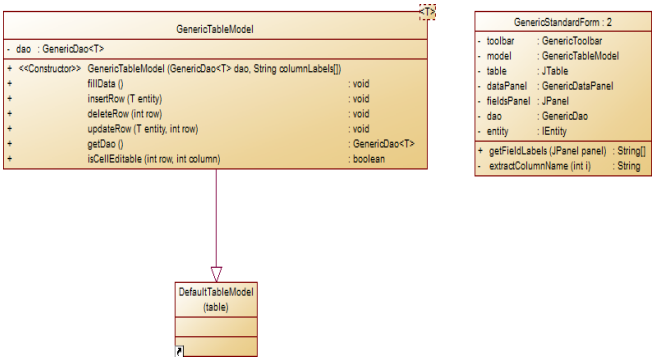


Figure 5. a) A sketch of a simple framework used for implementation of the administration subsystem application b) Details for *GenericStandardForm* and *GenericTableModel* classes

In order to support faster and easier development of the administration subsystem UI, we have developed a simple framework presented in Figure 5. Classes *GenericStandardForm* and *GenericTableModel* are used for data presentation and manipulation of all persistent classes in the middle-tier. Examples of administration subsystems forms developed with this framework are presented in Figure 6, 7, and 8. The framework code is available as a part of the Kroki administration subsystem at [11]

C. *Integration with Kroki*

After an enterprise application is sketched using the mockup editor and/or the lightweight UML editor, it can be executed using Kroki's desktop or web engine. If user rights and customized menus are not specified, the application has a default menu that provides activation of all developed application forms. This is suitable for the development phase and requirements elicitation based on

prototypes, but before deployment, the application must be customized to support every user role in the enterprise.

After launching, the administration subsystem is supplied with an XML file that has a list of developed resources (forms, reports, etc.) provided by the Kroki tool in the application repository (Figure 3). Kroki is "aware" of the administration subsystem's existence, but the reverse does not apply, in order to achieve a higher level of independence of the administration subsystem. Once the enterprise application is deployed, administrator should manage users and user rights using only the administration subsystem, with no need to access other Kroki's tools.

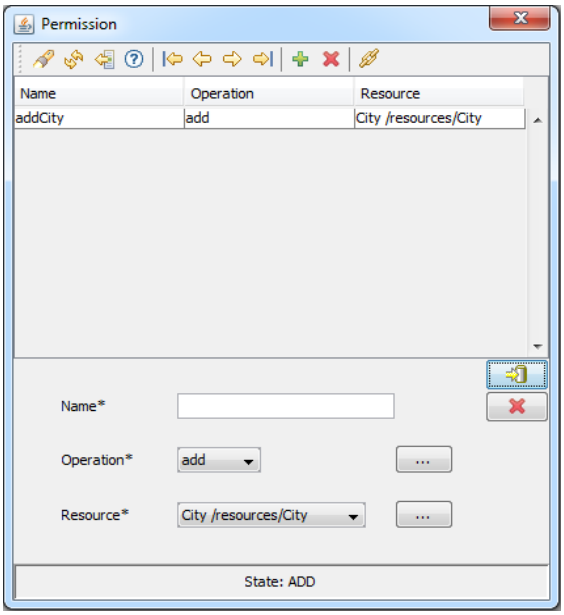


Figure 6. Permissions form

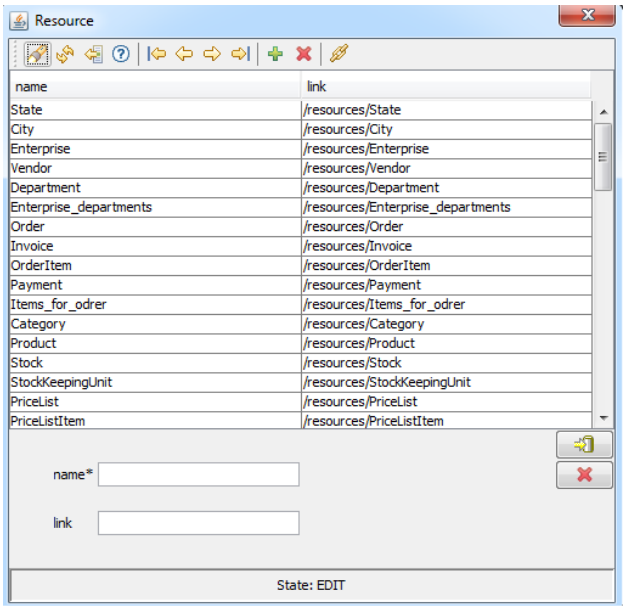


Figure 7. Resources form

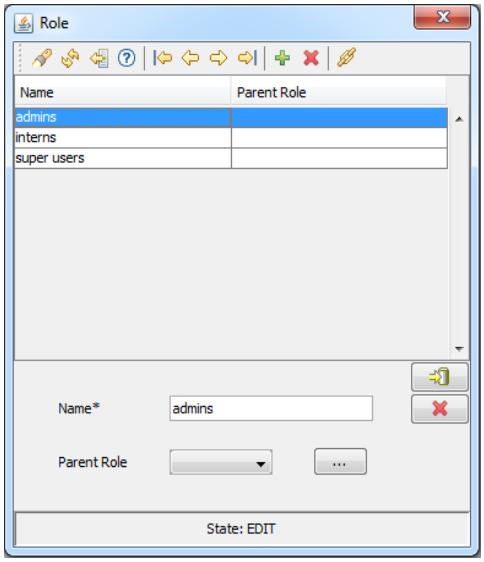


Figure 8. Roles form

The administration subsystem is storing user rights information and menu specifications in its own database. This information is provided to web and desktop engines in order to perform dynamic adjustments during run-time.

D. The Web and Desktop Engines

The web and desktop engines are performing dynamic adjustments using aspects. Kroki engines use aspect-oriented programming techniques to enable easier integration of cross-cutting concerns imposed by its tools.

The web engine is developed using restlets, so all of the web classes extend RestletResource class and are located in the resources package. Every resource class has `prepareContent` method that is invoked when a client request is sent to a particular resource and can be used to attach aspect functionalities. Restlet resources use map called `dataModel` to pass arbitrary data to HTML templates, so once attached to `prepareContent`, aspect can get access to the resource object and modify its `dataModel`. `DataModel` is wrapped into HTML elements using Freemarker templates (Figure 9).

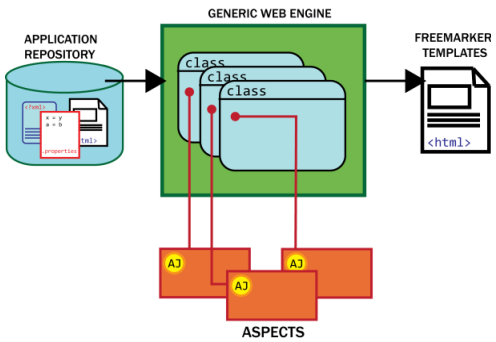


Figure 9. Web engine architecture

Listing 1 shows an aspect that modifies the main menu based on customization specified by the administration subsystem. It is activated after user has logged in and before menu is created in the application.

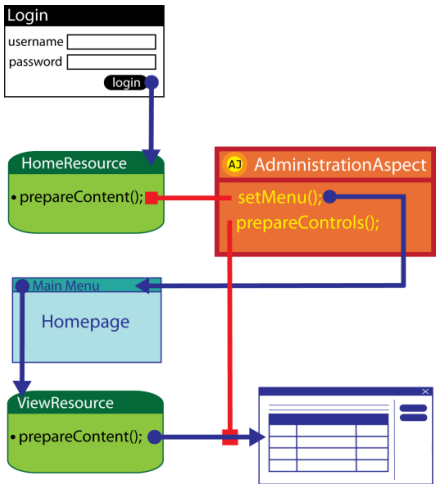


Figure 10 Aspects can change the content before pages are rendered

Since the generic web engine is designed as a single-page AJAX web application, once the user is logged in, the interaction takes place on the home page and restlet resource in charge of this page. So, in order to modify the main menu creation process, we need to attach our aspect to `prepareContent` method of `HomeResource` class. Freemarker template looks up main menu list by the name `main_menu`, so it will be the name by which we will put our modified menu into `dataModel` (Figure 10). Listing 1 represent basic steps described above.

```
public aspect MainMenuAspect {
//Create the pointcut that intercepts PrepareContent
//method in Home resource and obtain home resource object
public pointcut setMenu(HomeResource homeResource) :
        call(public void HomeResource.prepareContent()) &&
        this(homeResource);


        after (HomeResource homeResource):
                setMenu(homeResource) {
        User user = SessionAspect.getCurrentUser();
        //Obtain main menu list from AppCache
        ArrayList<AdaptMenu> menus =
                AppCache.getInstance().getMenuList();

        List<UserRoles> roles = …;
        //Obtain user roles through query
        if (roles.size() == 0) {
                //Put default main menu to data model
                homeResource.addToDataModel("menu", menus);
         else {
                //Retrieve and put modified
                //main menu to data model
                AdaptMenu modified_menu = …;
                homeResource.addToDataModel("menu",
                modified_menu);
                .
                .
                .
```

Listing 1. Aspect for menu loading

Similar activities are performed for the application forms in order to dynamically adjust their toolbar according to the specified user rights. More details about Kroki engines can be found in [10].

E. Menu Specification

The administration subsystem enables specification of customized menu for every user role. Custom menu allows personalization of users working environment.

Menus in administration subsystem are based on the composite design pattern (Figure 11).

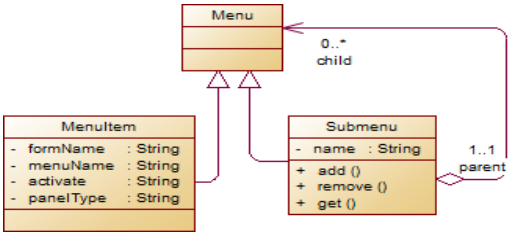


Figure 11. Menu structure within administration subsystem

The menus are stored in an XML file and deployed to the application repository. The tool for specification of menus is shown in Figure 12.

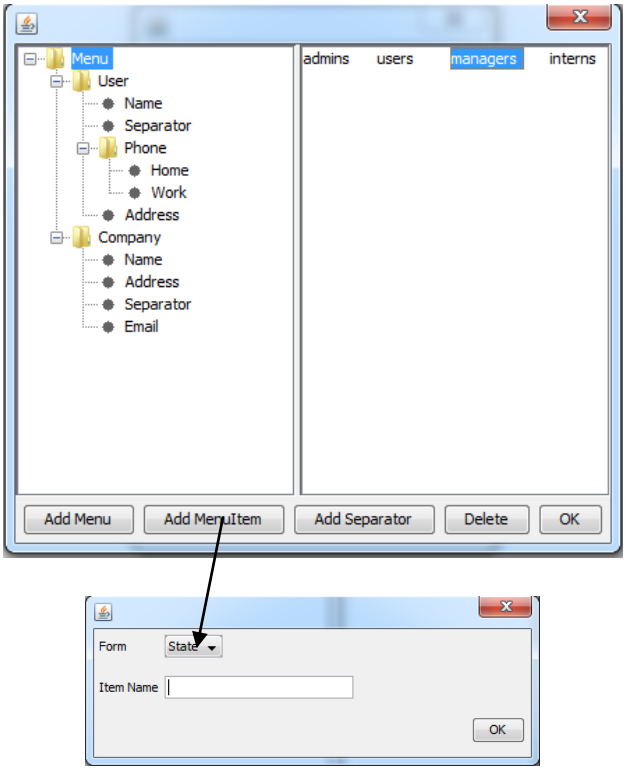


Figure 12. The tool for menus specification

IV. AN EXAMPLE

This section shows an example of the run-time adaptation of a business application according to the user roles.

Figure 13 and Figure 14 depict the same form in the web application. The depicted form is used by two groups of users. The first group of users is allowed to add, modify and remove data (see Figure 13), while the second group of users is allowed only to view data in that particular form (see Figure 14).

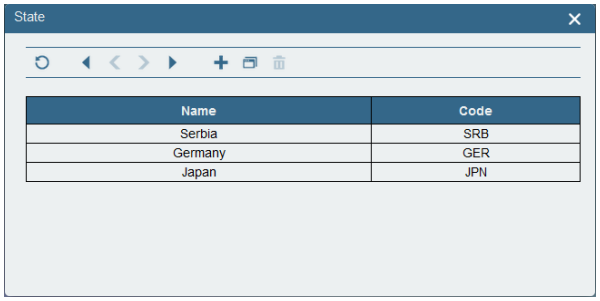


Figure 13. User form with add/modify/delete permissions

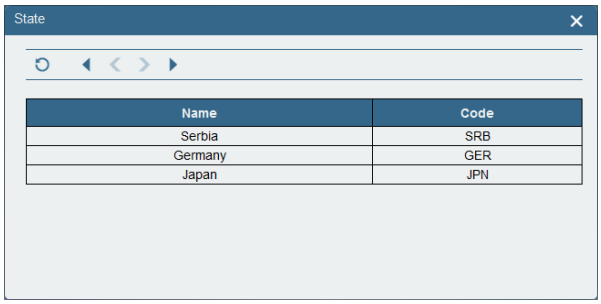


Figure 14. User form with view permission

Similarly Figures 15 and 16 show the same menu adapted for two different users. Available menu items are determined by user roles. In the case presented in Figure 15. the user is entitled to all submenus in the given menu, while in the case presented in Figure 16 the user is allowed to access only a subset of menu items.

Figure 15. Menu with all the submenus shown

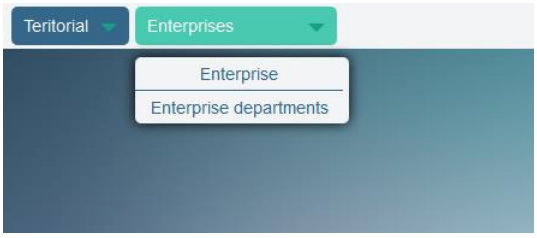


Figure 16. Menu with certain submenu restrictions

V. CONCLUSION

The paper presented the Kroki administration subsystem and generic AOP engines that enable enterprise application execution. The administration subsystem is based on RBAC standard for access control. RBAC-based systems enable users to perform previously authorized tasks, by dynamically adjusting the availability of actions.

Dynamic application adjustment is based on the application repository that contains configuration files and generic AOP engines. Configuration files are generated from the current Kroki model and the administration subsystem and used as a specification for application dynamic behavior.

Dynamic access to the user rights and their activation in appropriate moments is achieved with the use of AOP. AOP techniques enabled easier integration of cross-cutting concerns imposed by different Kroki tools and easier integration of the engines with hand-written code. Also, AOP allowed clear separation of duty and insight into expandable and easily readable code. Thanks to that, integration of the administration subsystem with Kroki performed seamlessly without the need to change the original tool.

REFERENCES

[1] R. Sandhu, E.J. Coyne, H.L. Feinstein, C.E. Youman. *Role-Based Access Control Models, IEEE Computer (IEEE Press) Vol. 29, Issue 2, pp. 38 - 47.*

[2] B. Trninić, G. Sladić, G. Milosavljević, B. Milosavljević, Z. Konjović, „*PolicyDSL: Towards Generic Access Control Management Based on a Policy Metamodel*“, SoMeT 2013, Budapest, Hungary.

[3] D. Kim, I. Ray, R. France, N. Li. *Modeling Role-Based Access Control Using Parameterized UML Model, FASE 2004. LNCS, vol. 2984, pp. 180-193*

[4] T. Lodderstedt, D. Basin, J. Doser, “SecureUML: A UML-Based Modeling Language for Model-Driven Security,” Proceedings of the 5th International Conference on The Unified Modeling Language, Dresden, Germany, September 30. - October 4. pp. 426-441, 2002.

[5] G. Milosavljević, M. Filipović, V. Marsenić, I. Dejanović. *Kroki: Interactive Development Of Business Application Based on Mockups*, Software Methodologies, Tools and Techniques 2013, Budapest, Hungary, pp. 235-242

[6] M. Filipović, *Adaptive Architecture Of Web Application Based on Aspects*, Master thesis, University of Novi Sad, 2011

[7] Kroki, www.kroki-mde.net

[8] Kroki demo, http://youtu.be/r2eQrl11bzA

[9] J. M. Rivero, J. Grigera, G. Rossi, E. Robles Luna, N. Koch, “Improving Agility in Model-Driven Web Engineering”, CAiSE Forum 2011, pp.163-170, 2011

[10] M. Filipović, S. Kaplar, R. Vaderna, Ž. Ivković, G. Milosavljević, I. Dejanović, Aspect-Oriented Engines for Kroki Models Execution, submitted to ICIST 2015, Kopaonik, Serbia

[11] Kroki Administration Subsystem Source, https://github.com/KROKIteam/KROKI-mockup-tool/tree/master/Kroki-Administration

RDF Stores Performance Test on Servers with Average Specification

Nikola Nikolić, Goran Savić, Milan Segedinac, Stevan Gostojić, Zora Konjović
University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia
{nikola.nikolic, savicg, milansegedinac, gostojic, ftn_zora}@uns.ac.rs

***Abstract* — The paper analysis the performances of different RDF stores on servers with average hardware specification. For this purpose, various tests have been performed on three RDF stores, namely Jena-Fuseki, Sesam-OWLIM and Virtuoso. Using different data sets and queries, the tests have measured CPU usage, heap memory consumption and execution time. Based on the results, for different application scenarios, an appropriate RDF store has been suggested.**

I. INTRODUCTION

Most applications that are based on semantic web store their data within RDF stores [1]. So far, various RDF stores have been developed providing different characteristics and performances. Depending on a particular application and its usage scenario, it must be decided which RDF store to use in order to satisfy both functional and non-functional requirements.

Many software applications do not contain complex functionalities implying that they do not store data sets larger than one million triplets. Executing queries over such smaller RDF stores doesn't require powerful and expensive servers. This paper is focused on such applications and is trying to propose an appropriate RDF store for non-expensive servers with average features.

The paper presents tests performed on three commonly used RDF stores. The tests have been ran within custom-made client application that executes various SPARQL queries to an RDF store. RDF stores have been set up on a server with average characteristics where RAM memory does not exceed 8 GB, CPU has up to 4 cores with disk drive space not larger than 500 GB.

Based on widely recognized RDF store ranking [2], we have chosen to test Virtuoso and Jena-Fuseki open source solutions and a free Lite version of commercial RDF store Sesame-OWLIM.

The following text has been organized as follows. The next section gives a short overview of similar performance tests. Section 3 describes our tests providing details about machine configuration used RDF stores, data sets and SPARQL queries. Measured parameters and testing procedure are also explained in this section. Results are presented and analysed in the section 4. Finally, the last section concludes the paper giving the future directions of this research.

The result of this testing is two-fold: on one hand, the proposal of the appropriate RDF store for a particular case of usage, and on the other hand comparison of performances of most commonly used RDF stores.

II. RELATED WORK

According to W3C list of references [3], several RDF benchmarks have been performed so far. These benchmarks mostly test large data sets by executing queries on powerful servers.

The most popular is Berlin SPARQL benchmark [4], which supports testing of several RDF stores, such as Sesame, Virtuoso, Jena-TDB, BigData and BigOwlim. It is based on a generic data set that is a part of an e-commerce use case. The data set contains a set of products, offered by different vendors and consumers, which post reviews about products. The performances have been measured on different size of data sets using various SPARQL queries. Data sets size varies from 10 million up to 150 billion triplets. Tests were performed on capable server machines worth up to ~70,000€ [5]. Given that such server highly exceeds hardware limitations set for our research, results of these tests cannot be used in the analysis conducted in this paper. Still, we have used the same testing procedure and SPARQL queries as Berlin SPARQL benchmark.

Another popular benchmark is SP^{2}Bench SPARQL benchmark performed on its own data sets, which are based on library scenarios. The benchmark uses smaller data sets consisting of up to a million triplets. In contrast to Berlin SPARQL benchmark, SP^{2}Bench evaluates performances of a single RDF store with variable RDF schemas [6].

The last benchmark we present in this paper is *Lehigh University Benchmark* (LUBM). The tests were performed on data sets that contain data on university publications. Successive batches of the same queries were used with some minor data variations. Such testing procedure does not represent real life scenarios where an application must response to a wide set of different queries.

The remaining SPARQL benchmarks listed in [3] do not cover all testing parameters that are relevant for our research (these parameters are described in Section 3, part E).

III. TEST

This section describes tests we have performed within this research.

A. Test machine

The tests have been run on a server with following features:

- Processor: Intel i5-3470 3.2GHz
- RAM: 8GB DDR3
- HDD: 500GB SATA3 7200rpm
- Operating system: Linux-Ubuntu 14.04.1 LTS 64-bit

In addition, one of the tested RDF stores works only with data loaded into working memory (*In-Memory Backend*). To make the results comparable, we have configured all data stores using *In-Memory* setup.

B. RDF stores

Following RDF stores have been used:

1. Virtuoso – Version 6.1.8,
2. Sesame-OWLIM – Version OWLIM-Lite 5.4.6486, based on Sesame 2.7.13, deployed on Apache-Tomcat 6.0.41.
3. Jena-Fuseki – Version 1.1.1, open source RDF store

Virtuoso [7] is free and open source software product which is one of the most commonly used RDF system worldwide. It is a universal hybrid server for handling data such as RDF triplets and XML documents. Using Virtuoso it is possible to combine SPARQL and SQL queries for handling RDF data.

Sesame-OWLIM is a commercial RDF store that offers a free version with limited features. This free version has been used in our study. It supports wide set of tools developed in Python and Java programming language. These tools provide increased RDF(S) functionalities. Sesame itself lacks OWL reasoner. To address this limitation, Sesame was upgraded with OWLIM third-party store [8] that adds missing functionalities. OWLIM belongs to newer generations of RDF stores which are made for more frequent data updating and increased concurrent access. Since beginning of 2014. the popularity of this RDF store has increased which drew our attention towards testing it. Free Lite version of OWLIM [9] is available with restricted features. One important constraint requires that data must be loaded in working memory.

Jena [10] is an open source software framework written in Java providing both storage and access of RDF data. It comes with its own OWL RDF graph reasoning component. It uses Fuseki SPARQL interface for accessing the abstract RDF graph model through HTTP protocol. Fuseki can be run as a stand-alone SPARQL server too.

C. Data sets

The tests use the same data sets as Berlin SPARQL benchmark. As mentioned, these data sets are taken from e-commerce domain containing sets of products that are classified by vendors and rated by reviewers. Data sets were programmatically generated in different sizes and representations depending on product count using BIBM (*Business Intelligence Benchmark*) generator [11].

Each data set was built from different class instances of vendors, producers, product offers, product types, product features, reviews, reviewers and their web pages. An example of a product class instance is shown in Listing 1.

```
dataFromProducer021:Product015
   rdfs:label "Dell Inspirion 3521";
   rdfs:comment "New machine";
   rdf:type ftn:Product;
   rdf:type ftn-inst:ProductType123;
   ftn:producer ftn-inst: Producer021;
   ftn:productFeature ftn-inst:ProductFeature456;
   ftn:productPropertyTextual1 "The best";
   ftn:productPropertyNumeric1 "17"^^xsd:Integer;
   dc:publisher dataFromProducer021:Producer021;
   dc:date "2015-01-07"^^xsd:date .
```

Listing 1. Product class instance

Each product is described with label, comment and product type. Product type defines different product features which are also described with label and comment. The product is produced by one or more vendors. A vendor is described with label, comment, web page URL and country URI. Offer is described with price, expiration and delivery date. Reviewers are described with name, e-mail address and nationality.

Table 1. shows the characterictics of the generated data sets. The data sets contain up to a million triplets. Table rows display number of class instances for the given number of expected triplets starting at 1K triplets and all the way up to 1M. Given that BIBM generator cannot generate the exact number of required triplets, the penultimate row display how many triplets have been generated. The last row presents the size of the file containing the data.

TABLE I.
CHARACTERISTICS OF DATA SETS

RDF triplets	1K	10K	100K	1M
Products	1	25	260	2848
Producers	1	1	6	61
Product Features	289	289	1954	4745
Product Types	7	7	37	151
Vendors	1	1	3	30
Offers	20	500	5200	56960
Reviewers	1	13	129	1451
Reviews	10	250	2600	28480
Exact RDF triplets	1844	10250	101817	1022446
File size (unzipped)	210.6kB	968.9kB	9.3MB	93.6MB

D. SPARQL queries

Data sets of all sizes have been tested using a combination of two groups of queries. First group gathers queries aimed on searching and navigation through the required product fragments. It includes 12 patterns [12], whereby the most important are:

1. Generic search for a given set of generic product properties.
2. More specific search for products with a given set of product properties.

3. Finding similar products of a given product.
4. Retrieving detailed information on several products.
5. Retrieving reviews for given products.
6. Getting background information about reviewers.
7. Retrieving offers for given products.
8. Checking information about vendors and their delivery conditions.

Second group of queries is designed to test independent analytical queries over the dataset. It includes 8 patterns [13]. These are:

1. The first 10 of most discussed product categories of products from a specific country which are based on number of reviews by reviewers from a certain country.
2. The first 10 products that are most similar to a specific product, rated by the count of features they have in common.
3. Products with the largest increase of interest (ratio of review counts) from one month to the next.
4. Feature with the highest ratio between price with that feature and price without that feature.
5. The most popular products of a specific product type for each country - by review count.
6. Reviewers who rated products by a specific Producer much higher than the average.
7. Products which are in first 1000 of most offered products of a certain product type that are not sold by vendors of a specific country
8. The top 10 cheapest vendors for a specific product type by the ratio of products below and above the average.

Both set of queries have been executed in random order using SPARQL protocol on chosen RDF stores.

E. Measuring parameters

RDF benchmarks explained in the Section 2 primarily measure these two variables:

- Time needed for loading and indexing of triplets
- Time needed for executing SPARQL queries

In our research we want to examine system performances in more details by measuring more parameters. Benchmark in this study is written in Java programming language. JvmTop [14] open source console application has been used for measuring the performances of RDF stores. For all running JVM (Java Virtual Machines) on a given system, JvmTop provides monitoring of following resources:

- Process ID
- Name of measured class
- Current heap memory usage depending on maximum allocated value
- Current non-heap memory usage depending on maximum allocated value
- CPU usage
- Percentage of garbage collector usage
- Number of infinite loops
- Number of created threads
- Thread state
- CPU usage by threads
- Number of blocking threads

All these variables were used to determine the system performance.

F. Method

Testing was done by implementing the following procedure for all data sets and RDF stores:

1. Load one data set in an RDF store
2. Execute first group of SPARQL queries
3. Execute second group of SPARQL queries
4. Save measured parameters results

Table 2. shows total number of executed queries per single data set. As mentioned, queries have been divided into two groups. Total of 15000 queries have been executed on each RDF store, where 10 000 queries belong to first group, while the remaining 5 000 queries belong to a second group of SPARQL queries.

TABLE II.
TOTAL EXECUTED SPARQL QUERIES

TOTAL EXECUTED QUERIES		
Data sets	Group 1	Group 2
1K, 10K, 100K	2500	1500
1M	2500	500

IV. RESULTS

This section presents the results of measuring four key parameters – CPU usage, heap memory usage, individual and total query execution time.

Table 3. shows CPU usage for all the tested RDF stores for a separate execution of queries on all data sets.

TABLE III.
CPU USAGE PER RDF STORE

RDF DB systems	Data sets	CPU USAGE [%]		
		MIN	MEAN	MAX
Jena-Fuseki	1K	0	31.32	103.47
	10K	0	27.19	105.85
	100K	0	28.66	400.00
	1M	0	21.54	100.00
Sesame-OWLIM Lite	1K	0	19.42	117.86
	10K	0	30.11	154.52
	100K	0	35.74	363.89
	1M	0	22.74	84.09
Virtuoso	1K	0	20.05	83.34
	10K	0	21.15	129.17
	100K	0	21.11	218.75
	1M	0	18.75	62.50

Results were shown in 3 columns representing minimum, maximum and mean value, expressed in percentages. We can notice that for Jena-Fuseki RDF store the maximum value reached as far as 400%. It corresponds to sum of usages of all loaded CPU cores. It can be noted that for data sets beginning at million triplets Virtuoso RDF store puts the lowest load on CPU.

Figures 1, 2, 3 and 4 show the charts of CPU usage in time during queries execution. The chart series are derived using 6th degree polynomial regression which by definition introduces certain error. It can be noticed that Sesame-OWLIM Lite store reached a slightly higher CPU load than the other two RDF stores.

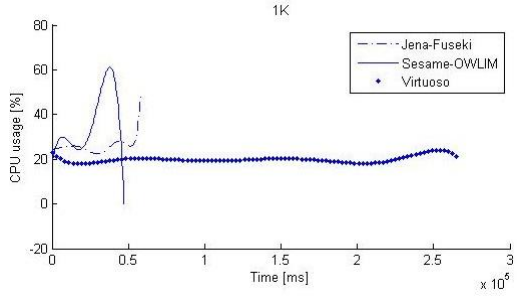


Figure 1. CPU usage for data set 1K

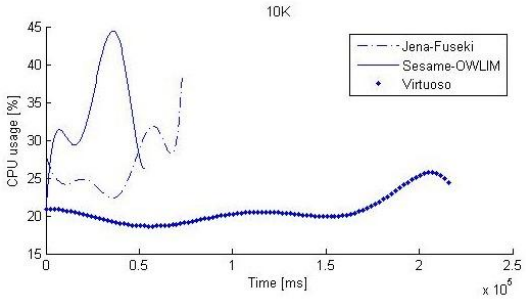


Figure 2. CPU usage for data set 10K

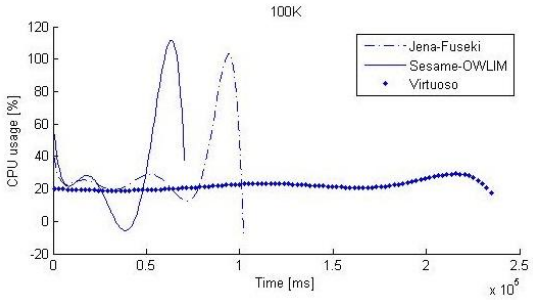


Figure 3. CPU usage for data set 100K

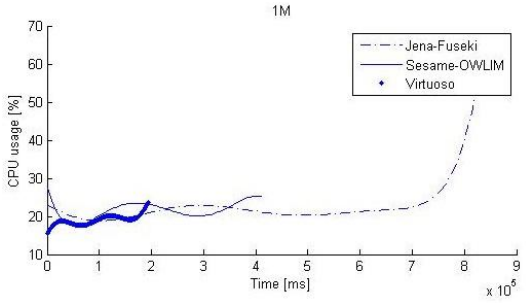


Figure 4. CPU usage for data set 1M

Given that all RDF stores use RAM memory to store the data, the measurement of heap memory usage has been necessary. Table 4. shows heap memory usage during queries execution over all data sets. If we analyse mean values, it can be noticed that Virtuoso occupies the least amount of heap memory on average.

TABLE IV.
HEAP USAGE PER RDF STORE

RDF DB systems	Data sets	HEAP USAGE [MB]		
		MIN	MEAN	MAX
Jena-Fuseki	1K	32.5	113.88	187
	10K	21.75	96.49	182
	100K	15.5	105.24	316.5
	1M	24	86.09	397
Sesame-OWLIM Lite	1K	9	69.93	175
	10K	17	104.61	178
	100K	16	108.12	292
	1M	25	86.46	396.5
Virtuoso	1K	10.25	51.69	171.5
	10K	17.25	64.82	175.5
	100K	13.5	70.31	176
	1M	20	67.67	173

Figures 5, 6, 7 and 8 show the charts of heap usage in time during queries execution.

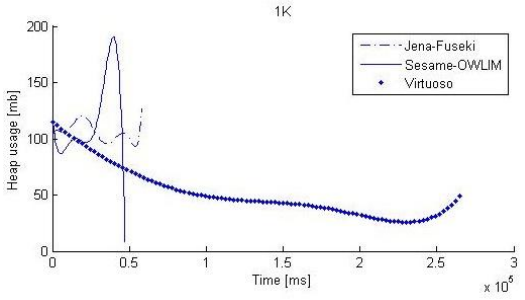


Figure 5. Heap usage for data set 1K

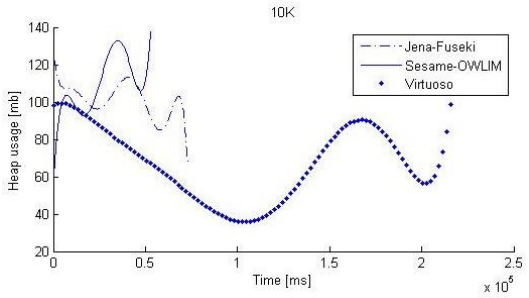


Figure 6. Heap usage for data set 10K

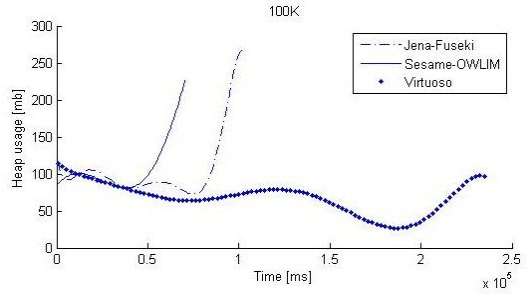


Figure 7. Heap usage for data set 100K

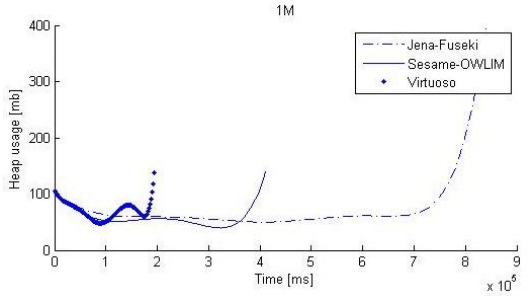


Figure 8. Heap usage for data set 1M

The third measured parameter is execution time of individual queries over all data sets. It is shown in Table 5.

TABLE V.
EXECUTE QUERY TIME PER RDF STORE

RDF DB systems	Data sets	EXECUTION TIME [ms]		
		MIN	MEAN	MAX
Jena-Fuseki	1K	3.08	12.08	1157.53
	10K	3.25	12.32	1114.64
	100K	3.01	20.41	1451.98
	1M	3.08	275.25	135709.74
Sesame-OWLIM Lite	1K	2.39	10.16	1081.49
	10K	2.31	9.10	1039.85
	100K	1.90	13.76	1057.26
	1M	2.30	136.28	65660.16
Virtuoso	1K	2.67	66.86	852.82
	10K	2.88	52.45	875.35
	100K	2.90	55.11	2300.94
	1M	3.04	58.59	3248.01

By observing mean values of execution queries, we can notice that Sesame-OWLIM Lite store takes considerably less time to execute SPARQL queries.

Figure 9, 10, 11 and 12 show the charts of individual query execution in time. Although Virtuoso has best performances in CPU and heap usage benchmarks, it doesn't show good results in execution time benchmark. However, for queries executed on data set of million triplets Virtuoso retains the lead.

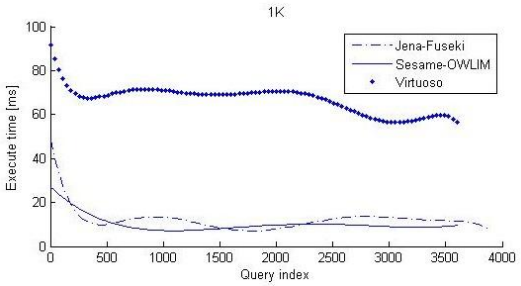


Figure 9. Execute time for data set of 1K triplets

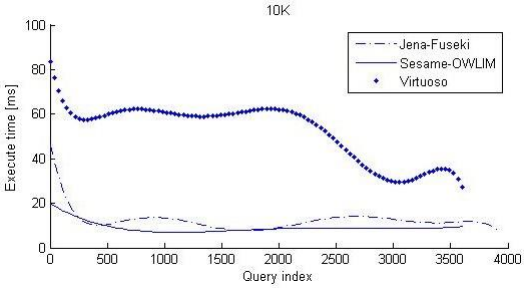


Figure 10. Execute time for data set of 10K triplets

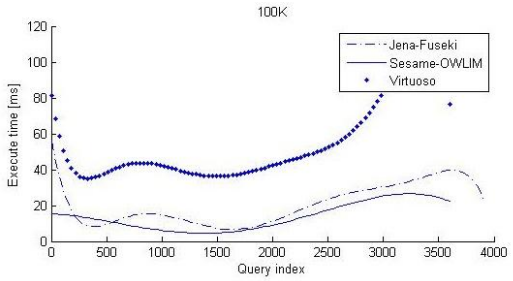


Figure 11. Execute time for data set of 100K triplets

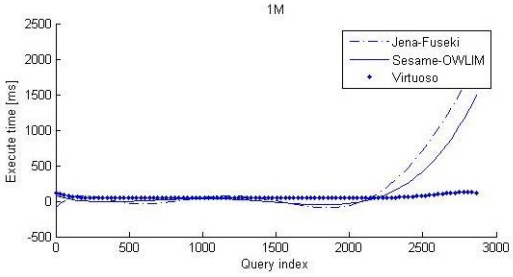


Figure 12. Execute time for data set of 1M triplets

Table 6 shows total execution time of both query groups on different data sets. Rows represent data sets, while columns represent RDF stores.

We can notice that Virtuoso store has notably better total execution time for data set of million triplets. Still, for data sets of less than one million triplets, Virtuoso store has by far worse total execution time. Sesame-OWLIM store has the best total execution time over data sets of less then million triplets.

TABLE VI.
TOTAL EXECUTION TIME

TOTAL EXECUTION TIME [s]			
Data sets	RDF DB systems		
	Jena-Fuseki	Sesame-OWLIM Lite	Virtuoso
1K	57.58	46.65	264.71
10K	72.36	52.07	215.27
100K	101.27	69.88	234.13
1M	838.89	410.01	192.91

In accordance with the results, we can determine which RDF store is the most appropriate for the particular scenario.

For data sets containing less than one million triplets we have shown that Virtuoso has better results in heap and CPU usage, but individual and total query execution time over such data sets is significantly greater than for Sesame-OWLIM Lite store.

For scenarios whose priority is the speed of query execution, we would recommend Sesame-OWLIM Lite store for data sets of less than one million triplets. However, for data sets greater than one million triplets, Virtuoso store is the most efficient regarding all parameters measured in this study.

V. CONCLUSION

In this paper we presented performance tests of commonly used RDF stores deployed on servers with average characteristics. The test results should facilitate the selection of RDF store for applications that do not work with data set larger than one million triplets. We tested Virtuoso and Jena-Fuseki as open source RDF stores, as well as Sesame-OWLIM as a free version of a commercial RDF store. The tests measured CPU and heap usage as well as time needed for query execution. The paper proposed recommendations for different scenarios, depending of the importance of the specific performance indicator.

Future research will be aimed on testing mentioned RDF data stores in native-storage mode where data is stored on a disk in contrast to the research presented in this paper where data is stored in RAM memory. In that case a commercial version of Sesame-OWLIM would be needed.

ACKNOWLEDGMENT

Results presented in this paper are part of the research conducted within the Grant No. III-47003, Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] W3C RDF, *http://www.w3.org/RDF/*

[2] DB-Engines Ranking, *http://db-engines.com/en/ranking_definition*

[3] W3C RDF store benchmarking, *http://www.w3.org/wiki/RdfStoreBenchmarking*.

[4] C. Bizer and A. Schultz. "The Berlin SPARQL Benchmark.", Int. J. Semantic Web Inf. Syst., 5(2):1–24, 2009.

[5] Berlin BSBM benchmark machine, *http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/results/V7/index.html#machine*.

[6] M. Schmidt, T. Hornung, G. Lausen, C. Pinkel. "SP^{2}Bench: A SPARQL performance benchmark.", ICDE, pages 222–233. IEEE, 2009.

[7] Virtuoso – OpenLink Software, *http://virtuoso.openlinksw.com*.

[8] A. Kiryakov, D. Ognyanov; D. Manov, OWLIM – a Pragmatic Semantic Repository for OWL, WISE 2005, 20 Nov, New York City, USA.

[9] Sesame-OWLIM Lite RDF store, *http://owlim.ontotext.com/display/OWLIMv54/OWLIM-Lite+Fact+Sheet*

[10] Apache Jena-Fuseki RDF store, *http://jena.apache.org/documentation/*

[11] BSBM generator, *http://sourceforge.net/projects/bibm/*

[12] SPARQL queries pattern – set 1., *http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/spec/ExploreUseCase/index.html#queriesTriple*

[13] SPARQL queries pattern – set 2., *http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/spec/BusinessIntelligenceUseCase/index.html#queriesTriple*

[14] JvmTop – Google code, *https://code.google.com/p/jvmtop/wiki/Documentation*

A Framework for ICT Support to Sustainable Mining - An Integral Approach

Nikola Zogović*, Sonja Dimitrijević*, Snežana Pantelić*, Dragan Stošić*
* University of Belgrade/Insitute Mihajlo Pupin, Belgrade, Serbia
{nikola.zogovic, sonja.dimitrijevic, snezana.pantelic, dragan.stosic} @pupin.rs

***Abstract*—Motivated by the facts that there is no one-fits-all sustainability assessment framework in mining and that support of information and communications technologies (ICT) to mining is focusing on specific mining process aspects we propose a cutting-edge ICT supported integral framework. The framework relays on multi-objective optimization theory, adaptive control theory, the mining process itself and modern ICT technologies, which position it in line with the concept of the Factory-of-Future. The framework should be able to provide data to all interested parties, e.g. to help top management in mining industry to make optimal decisions, to make available public data, to generate alarms in critical situations.**

I. INTRODUCTION

Mining is a fundamental human activity in the process of exploitation of natural ore resources [1]. Since the availability of ore highly affects existence of humans and progress of mankind, mining sustainability is of high importance. As we perceive nowadays, sustainable mining (SM) is leveraged by the five cornerstones [2]: economy, safety, environment pollution, production efficiency, and community.

The cornerstones are usually treated separately, with economy and production efficiency as the most important aspects, especially in developing societies, while safety, community and environment pollution aspects get their importance in high responsible societies mainly in developed countries.

Taking cornerstones for objectives, a multi-objective [3] approach to sustainable mining optimization can be performed, where all the objectives, intrinsically conflicting, are optimized simultaneously. Moreover, optimization of a mining system can be performed continuously to adapt the process to variable circumstances, such as weather conditions, current trading conditions on market regarding the subject ore, hazardous situations when system functionality can be reduced, etc.

Applying cutting-edge Information and Communications Technologies (ICT), including Internet of Things (IoT) [4, 5], Cyber Physical Systems (CPS) [6-8], Wireless Sensor Networks (WSNs) [9, 10], Cloud Computing (CC) [5, 11], Context-Aware Systems(CAS) [12], Data Mining (DM) [13], Machine Learning (ML) [14] and complex mathematical apparatus for multi-objective optimization [3, 15] to geology, mining engineering, machinery engineering, ecology, economy and finance expertise, the aim is to build a complex mining system that can be modelled, simulated or empirically studied in an integral and inter/multi-disciplinary approach with a goal of adaptive multi-objective optimization while satisfying sustainability condition. Such a system should enable top management of a mine corporation (and other interested sides) to have real-time information [16-18] and to make proper decisions.

Having the previous facts in mind we propose a framework for sustainably mining supported by the cutting-edge ICT that should bring mining to the Factory-of-Future concept, proclaimed by HORIZON 2020 – The European Commission program.

The paper is structured as follows. In section II we review the existing frameworks. In section III, we present the proposed framework with all its components. In section IV, we survey the existing ICT support to mining. In section V, we try to position our framework. Section VI concludes the paper.

II. RELATED WORK

Several ways for describing the term ‘‘Sustainability assessment framework’’ have been recognized. However, descriptions of the term are mostly ad-hoc. In general, “framework is a structure composed of components framed together to support something” [19, p.181]. In the case of sustainability assessment frameworks, these components are indicators/decision variables, conceptual models, principles, criteria, goals, and policies.

The number of frameworks that can be used to assess mining sustainability is on the rise. However, they are based on different approaches and may have different focuses. Moreover, their effectiveness is questionable [20] and requires more research.

Existing sustainability assessment and reporting frameworks primarily can be divided into two large groups [19]:

1. frameworks frequently used by mining companies [21-23]
2. frameworks proposed by analysts and academics, with a number or no implementation results though (e.g., seven questions to sustainability – 7QS [24], innovation and technology driven sustainability performance management framework – ITSPM [25], and Azapagic’s framework [26]

However, frequency of implementation is not a sufficient indicator of the effectiveness of a framework. Rare research studies focused on comparison and categorization of the frameworks (their attributes) confirm this claim [19, 27, 28]. These studies help clarify the positions of selected frameworks in the current theory and practice based on the analyzed variables such as temporal orientation, geographical focus, comprehensiveness / the number of decision variables, etc. As expected, the results of the comparative and other studies [29, 30] show that

there is no one-fits-all solution because of the complexity and variety of mining contexts. Moreover, they reveal some significant limitations of current frameworks that should be overcome by new approaches (e.g., weekly addressed geographical scope, predominantly retrospective temporal orientation, the neglected problem of scarcity of (proven) mineral reserves, etc.).

In an attempt to overcome obvious limitations of leading frameworks, new frameworks are continually being proposed (e.g., a systems-based framework for capturing the flows of materials across the globe [31], an indicator framework for measuring progress towards sustainability in the context of legacy mined land [32], etc.

III. FRAMEWORK

The proposed framework relays on multi-objective optimization theory, adaptive control theory, the mining process itself and modern ICT technologies.

A. Multi-Objective Optimization (MOO) Fundamentals

Let $\boldsymbol{z} = (z_1, \dots, z_n)$, $\boldsymbol{z} \in \mathbf{Z}$, where $\mathbf{Z}$ is the set of all feasible objectives' values, be the vector of objectives. Let $\boldsymbol{x} = (x_1, \dots, x_m)$, $\boldsymbol{x} \in \mathbf{X}$, where $\mathbf{X}$ is the set of all possible decision vector values, be the vector of decision (design) variables. Let $\boldsymbol{p} = (p_1, \dots, p_k)$, $\boldsymbol{p} \in \mathbf{P}$, where $\mathbf{P}$ is the set of all possible parameter vector values, be the vector of given parameters. Let $\mathbf{F}: (\mathbf{X}, \mathbf{P}) \rightarrow \mathbf{Z}$ be the mapping function from $\mathbf{X}$ and $\mathbf{P}$ spaces to $\mathbf{Z}$ space. Then $\mathbf{F} = (F_1, \dots, F_n)$, $\boldsymbol{z} = \mathbf{F}(\boldsymbol{x}, \boldsymbol{p})$, $z_1 = F_1(\boldsymbol{x}, \boldsymbol{p})$, …, $z_n = F_n(\boldsymbol{x}, \boldsymbol{p})$. Let $\mathbf{G}(\boldsymbol{x}, \boldsymbol{p}) = \mathbf{G}^*$ and $\mathbf{H}(\boldsymbol{x}, \boldsymbol{p}) \leq \mathbf{H}^*$ be the constraints given in equality and inequality forms, respectively.

For $\boldsymbol{p} = \mathbf{p}^*$, a point $\mathbf{x}^* \in \mathbf{X}$, is Pareto optimal (PO) iff there does not exist another point $\mathbf{x} \in \mathbf{X}$, such that $\mathbf{F}(\mathbf{x}, \mathbf{p}^*) \leq \mathbf{F}(\mathbf{x}^*, \mathbf{p}^*)$, and $F_i(\mathbf{x}, \mathbf{p}^*) \leq F_i(\mathbf{x}^*, \mathbf{p}^*), 1 \leq i \leq n$, for at least one function.

For $\boldsymbol{p} = \mathbf{p}^*$, a point $\mathbf{x}^* \in \mathbf{X}$, is weakly Pareto optimal (WPO) iff there does not exist another point $\mathbf{x} \in \mathbf{X}$, such that $\mathbf{F}(\mathbf{x}, \mathbf{p}^*) \leq \mathbf{F}(\mathbf{x}^*, \mathbf{p}^*)$.

In other words, a point is WPO if there is no other point that improves all of the objectives simultaneously. In contrast, a point is PO if there is no other point that improves at least one objective without detriment to another. PO points are also WPO, while WPO points are not PO.

The set of all PO points is called Pareto optimal set, Pareto front or Pareto frontier.

The task of MOO is to find all

$$\mathbf{x}^* = \operatorname{argmax}_{\substack{\mathbf{G}(\boldsymbol{x}, \mathbf{p}^*) = \mathbf{G}^*, \\ \mathbf{H}(\boldsymbol{x}, \mathbf{p}^*) \leq \mathbf{H}^*}} \mathbf{F}(\boldsymbol{x}, \mathbf{p}^*).$$

B. Adaptive Control Theory Fundamentals

Control theory is a useful tool for control of system dynamics, another way around, for control of system transition from some state to the desired state. In addition, adaptive control should help system keep desired state, in the case of variable system parameters change, where the desired system state can also vary with the parameters change. Basics of the adaptive control are given in Fig. 1, e.g. see [33, 34].

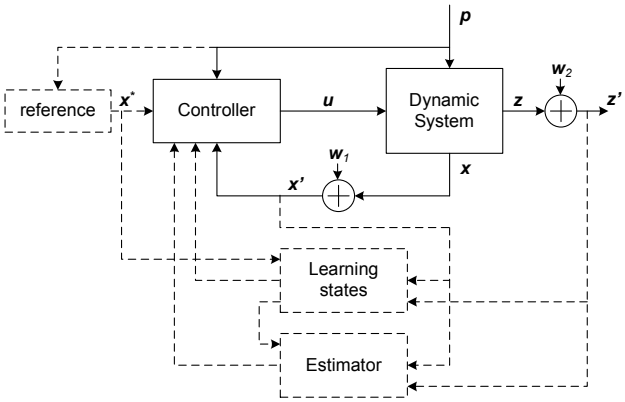


Figure 1. Adaptive control - basic diagram

The system state $\boldsymbol{x}$ and output $\boldsymbol{z}$ can be obtained by some measurement that introduces measurement errors $\boldsymbol{w_1}, \boldsymbol{w_2}$, respectively. Controller calculates optimal control vector $\boldsymbol{u}$, using all available data. Dashed lines mean optional. The way information regarding desired system state is obtained and specific adaptive control system architecture is chosen, depends on a particular controlled dynamic system.

Combining MOO and adaptive control, in the case the given parameters are not fixed, if they change independently of our control, for any change of $\boldsymbol{p}$, from $\mathbf{p}_1^*$ to $\mathbf{p}_2^*$ or some hazardous situation or any other system change reason, dynamic MOO [35] control mechanism, whichever the dynamism origins are [36], should adapt moving optimal point from $\mathbf{x}_1^*$ to $\mathbf{x}_2^*$, keeping transition slight, without running system out of function region or through the undesirable states.

Once having system state data and observed objectives, it is easy to generate alarms on critical system states and present data to interested parties.

C. Objective Space Analysis

Objective space can be qualitatively defined based on the five SM cornerstones [2], assigning, at least one, objective to each cornerstone.

Economy shows the difference of revenue and cost and the goal is to maximize benefit to all stakeholders [37]. Economy depends on many things and it is usually seen as the ultimate goal. Profitability analysis in mining starts from estimation of total ore amount in a mine region and target minerals' concentration in ore by geologists. Economy answers sustainability problem in mining by introduction of circular economy [38].

Production efficiency [39, 40] gives the ratio of produced goods (tones of target mineral or ore) and consumed resources (machinery, fuel, employees, etc.) in production process.

Safety [41, 42] can be expressed as the number of injuries in time unit, relatively to the number of employees or per the amount of produced goods or absolutely. Depending on consequences the injuries range from light, resulting in lost work time [43] to fatal, when human lives are lost.

Environment including air, water and soil is inevitably polluted in mining process [44-48]. Pollution can be expressed as the concentration distribution of pollutant. A decision support system for environmental reclamation of an open-pit mine is presented in [49]. Regulations,

monitoring and control of dust in mineral industries are surveyed in [50].

Community (local, state level, regional or even global) relates to mining twofold [51-53]. The first relation is directed from community to the mine corporation: does the mine corporation sufficiently attract individuals to keep the mining process and is the community sufficiently strong (numerous) to support the mining process? The second relation concerns how the mining contributes to the community wellbeing. The contribution is direct through the investments of the mine corporation to the community budget [54] and indirect, through the attracted investments to the community due to the mine existence. Contribution can be expressed as the fraction of the community GDP.

D. MOO Constraints

Having five objectives, sustainability [55, cf. ch. I, sec. B-I] can be expressed as the integral condition that all or some (the more the better) objective functions are non-decreasing[1] in time [19] or, having in mind that man is the measure of all things (Protagoras), it can be expressed as the non-decreasing contribution to the development of the universal human rights [48, 56] in community, in terms of health, education and living standards.

E. Design Space Analysis

We use two levels process identification to define decision space. At the first level, we consider a life cycle of a mine through all its phases. At the second level, we focus on the phase of ore exploitation as the main phase in mining and consider all the sub-phases in general, without details regarding specific ore exploitation or type of a mine.

The main phases in life cycle of a mine are shown in Fig. 2. In strategic level land exploration and ore detection phase analysis of spatial ore distribution within ground as well as evaluation of ore fraction are performed. Mine establishment phase relates to all activities needed for physical placement of a mine. Ore exploitation is the main phase and more details are given in the next section. After the exploitation phase, the mine should be closed while all the equipment and accompanying facilities should be removed. Finally, ground should be rehabilitated and the state before mine restored to keep environmental balance.

Figure 2. Lifecycle of a mine

[1] For mathematical clarity pollution and safety should be considered as the inverse functions

Ore exploitation phase consists of several sub-phases including the first and the second phases at tactical level. The phases are:

3. Tactical level land exploration and ore detection
4. Ore extraction method selection and the technique parameters setting, extraction schedule and location plan
5. Ore transportation technique selection and the technique parameters setting
6. Storage type of equipment selection
7. Equipment maintenance
8. Market, proper time and ore amount to be traded selection (trading)
9. Revenue management
10. Safety methods selection and safety plan (safety provision)
11. Environment protection methods selection and protection plan establishment

Ore exploitation sub-phases can be grouped into three groups: economy related, background processes and ore obtaining processes. Diagram of grouped sub-phases are shown in Fig. 3.

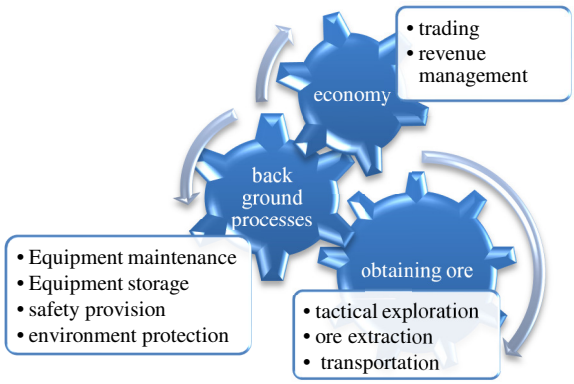


Figure 3. A mine in exploitation phase

We give some qualitative examples of decision variables arising from exploitation sub-phases:

1. Machinery fuel consumption
2. The set of engaged machines
3. The engagement schedule
4. The preventive maintenance schedule
5. Geographical plan of mining
6. The set of employees
7. Water-curtain air-clean system activation conditions and schedule
8. Blasting [57] or drilling schedule and space distribution
9. Design parameters of water-jet slotting system [58]

IV. EXISTING ICT SUPPORT TO MINING

There are a number of existing applications of ICT in mining. Here we list some examples following the states of control cycle given in Fig. 4.

For *data collection* in mining ICT provides several concepts with IoT, WSN, and CPS the most promising. IoT as a concept for information and control systems that are employed by mining industries is introduced in [4].

In [5] IoT and CC based system is employed to improve mine tailings dam safety. It is accomplished with the abilities of real-time monitoring of the saturated line, impounded water level and the dam deformation with the objective to provide pre-alarm information automatically and remotely in all weather conditions.

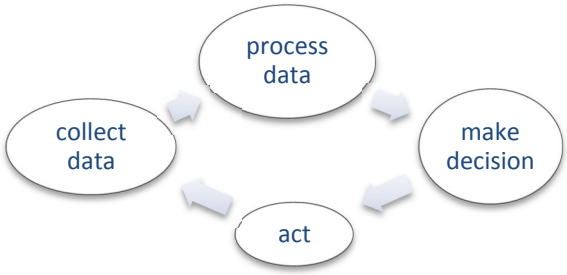


Figure 4. Control cycle

Locata [18] is the CPS for reliable positioning and navigation, invented by Locata Corporation, a Canberra-based company. Locata is a terrestrial high accuracy positioning system that can augment Global Navigation Satellite System (GNSS) with extra-terrestrial signals to permit cm-level positioning accuracy even when there are insufficient GNSS satellite signals.

A CPS targeting safety and accident rescue occurred in underground coal mine is proposed in [6]. The CPS senses surrounding, estimates environment parameters and based on personnel location generates warnings.

A CPS for unconfined compressive strength of rocks surrounding access tunnels in long-wall coal mining estimation can be designed based on Mamdani fuzzy model, proposed in [7]. The CPS should use Schmidt hardness, density, or porosity as the input parameters.

WSN based CPS for fire detection, location and spreading direction determination in a Bord-and-Pillar underground coal mine with capability to stop fire spreading is proposed in [9].

Structure-Aware Self-Adaptive WSN system able to rapidly detect structure variations caused by underground collapses and robust mechanism for efficiently handling queries under the instable circumstances is proposed in [10].

Data processing in mining can be efficiently supported by ICT through CC, DM and ML services. Some existing applications follow.

A new approach to implementation of a geo-information environment for the mining geo-information science problem solution using CC technologies is analyzed in [11].

A new emission model namely computer model for the construction of model-ready emission inventories (MOSESS) which is being used to compile high-resolution emission inventories or improve existing ones, utilizing complex GIS techniques is proposed in [13]. MOSESS, as well as the other similar models, such as AERMOD [59], can be efficiently realized using CC and DM technologies.

A remote sensing-based methodology for quantifying the impact of surface mining activity and reclamation from a watershed to local scale is proposed in [14]. The method is based on a Support Vector Machines (SVMs) classifier combined with multi-temporal change detection of Landsat Thematic Mapper imagery.

Decision making in mining can be supported by ICT in almost all phases and sub-phases. Several existing examples follow.

Context-aware intelligent service system that can be used to provide the most appropriate information services to miners according to their real-time situation, enabling self-safety control is proposed in [12], The system address questions regarding modeling the served miners' context, provision of the information service that meets miners' customized demands and verification of service invocation availability.

A framework for modeling interaction between a CPS and its environment is proposed in [60]. The lack of the interaction modeling can result in invalid worst-case estimation of system's safety and reliability. The framework supports simulation-based risk analysis of an initiating event such as equipment failure or flooding.

The decision support for optimal reclamation method using an AHP-based model for coal production in an open-pit coal mine located at Seyitomer region in Turkey is proposed in [49].

To help decision makers to optimally achieve objectives such as production lines reliability, maintaining costs, and system failure and downtime, employing preventive maintenance scheduling of complex manufacturing equipment the multi-objective optimal approaches are proposed in [61, 62].

An example showing how ICT can support *action* state of control cycle is a CPS for utilizing optimal switching control and a variable speed drive based optimal control. The CPS is proposed in [8] with the objective to improve the energy efficiency of belt conveyor systems at the operational level, under the constraints of time-of-use tariff, ramp rate of belt speed and other system parameters.

The need for an integral concept such as the proposed framework is validated by the previous examples, which tackle the framework just in some particular aspects.

V. POSITIONING OUR FRAMEWORK

We used the main attributes of sustainability assessment and reporting frameworks proposed in [19] to position our framework in the framework design space. The corresponding attributes are: temporal orientation, spatial or geographical focus, comprehensiveness, integration, scale and scope considerations. We also included ICT considerations in the set of attributes, as we find this attribute very important and unjustifiably neglected for comparing and positioning sustainability assessment frameworks. The proposed framework was described based on the given attributes in Table 1.

TABLE 1. POSITIONING THE FRAMEWORK IN THE FRAMEWORK DESIGN SPACE

Attribute	Position
Temporal orientation	The framework is aimed to be both retrospective and prospective. Past year data and indicators based on them are necessary for considering the future implications of mining operations to sustainability. The mine life cycle is taken into consideration as explained in Section III-E.

Geographical focus	The framework should enable clear specification of spatial boundaries and criteria for aggregating data from multiple sites, if necessary. This would allow for broader corporation, i.e. organizational-centered perspective.
Comprehensiveness	The framework is conceptualized as adaptive to change of parameters. Moreover, it should be scalable to SME and large companies, to private and public companies. Consequently, the proposed framework will incorporate a wider set of decision variables from which a subset adequate for a particular domain could be selected. The decision variables are focused on the defined cornerstones, but not limited to.
Integration	The framework especially targets the decision variables that have conflicted impact on the objectives. These include both integrated and non-integrated variables (see Section III-C).
Scale and scope considerations	Scale considerations (i.e. how performance varies across local, regional, national and global scales) are largely overlooked by current frameworks. Since this is an attribute hard to achieve in target economy, it is not in focus of the proposed framework. Scope considerations stem from the five cornerstones [2]. However, specific decision variables, protocols, policies and principles are expected to be drawn from different bodies of literature.
ICT considerations	The framework seeks to improve decision management thanks to historical, real-time and simulation information. To be able to achieve this goal, ICT support is of crucial importance (as explained in Section I). However, specific ICT support in an implementation environment should depend on needs and interests of a mine corporation.

In this light, we believe that what particularly distinguishes the proposed framework is scalability (comprehensiveness), importance given to prospective temporal orientation, as well as strong focus on decision management that requires modern ICT support.

VI. CONCLUSION

We propose an integral framework for sustainable mining taking into account the five SM cornerstones: economy, safety, environment pollution, production efficiency, and community. It relays on multi-objective optimization theory, adaptive control theory, the mining process itself and modern ICT technologies.

Following multi-objective optimization concept and using the mining process, we try to qualify objective space, decision/design space and constraints. Using adaptive control theory we introduce dynamic component of the mining process and try to adapt the system to any system disturbance. We evaluate how the current mining employs ICT and show that an ample space for ICT support exists.

Compared with the other sustainability frameworks we find that scalability (comprehensiveness), importance given to prospective temporal orientation, as well as strong focus on decision management that requires cutting-edge ICT support distinguishes it from the rest.

Moreover, further development of the proposed framework could bring mining to the Factory-of-Future concept.

ACKNOWLEDGMENT

The Ministry of Education, Science and Technological Development of Republic of Serbia supported the work by grants TR-32051 and TR-35030.

REFERENCES

[1] Behrens, Arno, et al. "The material basis of the global economy: Worldwide patterns of natural resource extraction and their implications for sustainable resource use policies." Elsevier - *Ecological Economics* 64.2 (2007): 444-453.

[2] Laurence, David. "Establishing a sustainable mining operation: an overview." *Journal of Cleaner Production* 19.2 (2011): 278-284.

[3] Marler, R. Timothy, and Jasbir S. Arora. "Survey of multi-objective optimization methods for engineering." *Structural and multidisciplinary optimization* 26.6 (2004): 369-395.

[4] Oldřich, Kodym, Danel Roman, and Kohut Vladimír. "Mining Production Information and Visualization Systems Based on Internet of Things." *Mine Planning and Equipment Selection*. Springer International Publishing, 2014. 911-920.

[5] Sun, Enji, Xingkai Zhang, and Zhongxue Li. "The internet of things (IOT) and cloud computing (CC) based tailings dam monitoring and pre-alarm system in mines." *Safety science* 50.4 (2012): 811-815.

[6] Sun, Yanjing, et al. "Model of Cyber-Physical Systems for Underground Coal Mine." *Green Communications and Networks*. Springer Netherlands, 2012. 3-11.

[7] Rezaei, Mohammad, Abbas Majdi, and Masoud Monjezi. "An intelligent approach to predict unconfined compressive strength of rock surrounding access tunnels in longwall coal mining." *Neural Computing and Applications* 24.1 (2014): 233-241.

[8] Zhang, Shirong, and Xiaohua Xia. "Optimal control of operation efficiency of belt conveyor systems." *Applied Energy* 87.6 (2010): 1929-1937.

[9] Bhattacharjee, Sudipta, et al. "Wireless sensor network-based fire detection, alarming, monitoring and prevention system for Bord-and-Pillar coal mines." *Journal of Systems and Software* 85.3 (2012): 571-581.

[10] Li, Mo, and Yunhao Liu. "Underground coal mine monitoring with wireless sensor networks." *ACM Transactions on Sensor Networks (TOSN)* 5.2 (2009): 10.

[11] Bychkov, I. V., V. N. Oparin, and V. P. Potapov. "Cloud technologies in mining geoinformation science." *Journal of Mining Science* 50.1 (2014): 142-154.

[12] Xue, Xiao, Jing-kun Chang, and Zhi-zhong Liu. "Context-aware intelligent service system for coal mine industry." Elsevier - *Computers in Industry* 65.2 (2014): 291-305.

[13] Markakis, Konstantinos, et al. "MOSESS: A New Emission Model for the Compilation of Model-Ready Emission Inventories—Application in a Coal Mining Area in Northern Greece." *Environmental Modeling & Assessment* 18.5 (2013): 509-521.

[14] Petropoulos, George P., Panagiotis Partsinevelos, and Zinovia Mitraka. "Change detection of surface mining activity and reclamation based on a machine learning approach of multi-temporal Landsat TM imagery." *Geocarto International* 28.4 (2013): 323-342.

[15] Miettinen, K. M., and Non-Linear Multi-Objective Optimization. "Kluwer Academic Publisher." (1999).

[16] Benndorf, Jörg. "Moving towards Real-Time Management of Mineral Reserves–A Geostatistical and Mine Optimization Closed-Loop Framework." *Mine Planning and Equipment Selection*. Springer International Publishing, 2014. 989-999.

[17] Benndorf, Jörg. "Real-time mineral resource models: Approaches for the integration of online production data." *Freiberger Forschungsforum 2014, 15. Geokinematischer Tag, Freiberg, Germany, 15-16 May 2014*. TU Bergakademie Freiberg, 2014.

[18] Rizos, Chris, et al. "Open cut mine machinery automation: Going beyond GNSS with Locata." *Proc. 2nd Int. Future Mining Conf., Sydney, Australia*. 2011.

[19] Fonseca, Alberto, Mary Louise McAllister, and Patricia Fitzpatrick. "Measuring what? A comparative anatomy of five

mining sustainability frameworks." *Minerals Engineering* 46 (2013): 180-186.

[20] Petrie, J., B. Cohen, and M. Stewart. "Decision support frameworks and metrics for sustainable development of minerals and metals." *Clean Technologies and Environmental Policy* 9.2 (2007): 133-145.

[21] Sustainability Reporting Guidelines & Mining and Metals Sector Supplement. Global Reporting Initiative (GRI), Amsterdam, 2010.

[22] Sustainability Reporting Guidelines – Version 3.1. GRI, Amsterdam. 2011.

[23] Towards Sustainable Mining: Progress Report 2011. Mining Association of Canada (MAC), 2012.

[24] America, MMSD North. "Seven questions to sustainability: how to assess the contribution of mining and minerals activities." *International Institute for Sustainable Development (IISD), Winnipeg, Manitoba* (2002).

[25] Basu, Arun J., and Uday Kumar. "Innovation and technology driven sustainability performance management framework (ITSPM) for the mining and minerals sector." *International Journal of Surface Mining* 18.2 (2004): 135-149.

[26] Azapagic, Adisa. "Developing a framework for sustainable development indicators for the mining and minerals industry." *Journal of cleaner production* 12.6 (2004): 639-662.

[27] Ness, Barry, et al. "Categorising tools for sustainability assessment." *Ecological economics* 60.3 (2007): 498-508.

[28] Hacking, Theo, and Peter Guthrie. "A framework for clarifying the meaning of Triple Bottom-Line, Integrated, and Sustainability Assessment." *Environmental Impact Assessment Review* 28.2 (2008): 73-89.

[29] Hodge, R. Anthony, and P. Eng. "Tracking progress toward sustainability: linking the power of measurement and story." Transactions – Society for Mining Metallurgy and Exploration Incorporated 320 (2007): 63.

[30] Pintér, László, et al. "Bellagio STAMP: Principles for sustainability assessment and measurement." *Ecological Indicators* 17 (2012): 20-28.

[31] Fiksel, Joseph. "A framework for sustainable materials management." *JOM* 58.8 (2006): 15-22.

[32] Worrall, Rhys, et al. "Towards a sustainability criteria and indicators framework for legacy mine land." *Journal of Cleaner Production* 17.16 (2009): 1426-1434.

[33] Krstic, Miroslav, Ioannis Kanellakopoulos, and Peter V. Kokotovic. *Nonlinear and adaptive control design*. Wiley, 1995.

[34] Åström, Karl J., and Björn Wittenmark. *Adaptive control*. Courier Dover Publications, 2013.

[35] Farina, Marco, Kalyanmoy Deb, and Paolo Amato. "Dynamic multiobjective optimization problems: test cases, approximations, and applications." *Evolutionary Computation, IEEE Transactions on* 8.5 (2004): 425-442.

[36] Tantar, Emilia, A. Tantar, and Pascal Bouvry. "On dynamic multi-objective optimization, classification and performance measures." *Evolutionary Computation (CEC), 2011 IEEE Congress on*. IEEE, 2011.

[37] Huang, Xueli, and Ian Austin. Chinese investment in Australia: Unique insights from the mining industry. Palgrave Macmillan, 2011.

[38] Mathews, John A., and Hao Tan. "Progress toward a circular economy in China." *Journal of Industrial Ecology* 15.3 (2011): 435-457.

[39] Li, Wei Guang, Chao Li, and Ke Li Chen. "Study on Technology of Mechanized Mining in Extremely Thin Coal Seam." *Advanced Materials Research*. Vol. 962. 2014.

[40] Thatcher, Matt E., and Jim R. Oliver. "The impact of technology investments on a firm's production efficiency, product quality, and productivity." *Journal of Management Information Systems* 18.2 (2001): 17-46.

[41] Burgess-Limerick, Robin, et al. "EDEEP—An Innovative Process for Improving the Safety of Mining Equipment." *Minerals* 2.4 (2012): 272-282.

[42] Shandro, Janis A., et al. "Perspectives on community health issues and the mining boom–bust cycle." Elsevier, *Resources Policy* 36.2 (2011): 178-186.

[43] Coleman, Patrick J., and John C. Kerkering. "Measuring mining safety with injury statistics: Lost workdays as indicators of risk." *Journal of safety research* 38.5 (2007): 523-533.

[44] Monjezi, M., et al. "Environmental impact assessment of open pit mining in Iran." *Environmental geology* 58.1 (2009): 205-216.

[45] Huertas, José I., et al. "Air quality impact assessment of multiple open pit coal mines in northern Colombia." *Journal of environmental management* 93.1 (2012): 121-129.

[46] Huertas, Jose I., Dumar A. Camacho, and Maria E. Huertas. "Standardized emissions inventory methodology for open-pit mining areas." *Environmental Science and Pollution Research* 19.7 (2012): 2784-2794.

[47] Chaulya, Swades-Kumar. "Air quality status of an open pit mining area in India." *Environmental monitoring and assessment* 105.1-3 (2005): 369-389.

[48] Kemp, Deanna, et al. "Mining, water and human rights: making the connection." *Journal of Cleaner Production* 18.15 (2010): 1553-1562.

[49] Bascetin, A. "A decision support system using analytical hierarchy process (AHP) for the optimal environmental reclamation of an open-pit mine." *Environmental Geology* 52.4 (2007): 663-672.

[50] Petavratzi, E., S. Kingman, and I. Lowndes. "Particulates from mining operations: A review of sources, effects and regulations." *Minerals Engineering* 18.12 (2005): 1183-1199.

[51] Kemp, Deanna. "Community relations in the global mining industry: exploring the internal dimensions of externally orientated work.", Wiley - *Corporate Social Responsibility and Environmental Management* 17.1 (2010): 1-14.

[52] Garvin, Theresa, et al. "Community–company relations in gold mining in Ghana." Elsevier -*Journal of environmental management* 90.1 (2009): 571-586.

[53] Lockie, Stewart, et al. "Coal mining and the resource community cycle: a longitudinal assessment of the social impacts of the Coppabella coal mine." *Environmental Impact Assessment Review* 29.5 (2009): 330-339.

[54] McIlmoil, R., et al. "Coal and renewables in Central Appalachia: The impact of coal on the West Virginia state budget", Charleston: West Virginia Center on Budget and Policy, 2010.

[55] "Measuring sustainable development", Report of the Joint UNECE/OECD/Eurostat Work Group on Statistics for Sustainable Development, UN, NY and Geneva, 2008

[56] "Measuring Democracy and Democratic Governance in a post-2015 Development Framework", UNDP, discussion paper, August 2012. Available at URL: www.icnl.org/research/resources/post_2015_development_adgenda/index.html

[57] Singh, T. N., and Virendra Singh. "An intelligent approach to prediction and control ground vibration in mines." *Geotechnical & Geological Engineering* 23.3 (2005): 249-262.

[58] Lu, Tingkan, et al. "Improvement of methane drainage in high gassy coal seam using waterjet technique.", Elsevier, *International Journal of Coal Geology* 79.1 (2009): 40-48.

[59] Reference available at URL: www.epa.gov/scram001/7thconf/aermod/aermod_mfd.pdf

[60] Sierla, Seppo, et al. "Common cause failure analysis of cyber–physical systems situated in constructed environments." *Research in Engineering Design* 24.4 (2013): 375-394.

[61] Ebrahimipour, V., A. Najjarbashi, and M. Sheikhalishahi. "Multi-objective modeling for preventive maintenance scheduling in a multiple production line." *Journal of Intelligent Manufacturing* (2013): 1-12.

[62] Berrichi, Ali, et al. "Bi-objective ant colony optimization approach to optimize production and maintenance scheduling." *Computers & Operations Research* 37.9 (2010): 1584-1596.

High level design of architecture for software reliability management of Power Supply Company Jugoistok

Aleksandar Dimov*, Nikola Davidović**, Leonid Stoimenov**
* Faculty of Mathematics and Informatics, University of Sofia, Bulgaria
aldi@fmi.uni-sofia.bg
** Faculty of Electronic Engineering, University of Niš, Niš, Serbia
nikola.davidovic@elfak.ni.ac.rs, leonid.stoimenov@elfak.ni.ac.rs

***Abstract*—Power supply companies require IT support and infrastructure at heterogeneous levels that vary from supervisory control to low level management of field devices. On the other side quality of software for power supply systems is important characteristic to consider, but it is difficult to be measured and reason about in formal way. This is mostly due to the fact, that for such systems testing is not always applicable into real-world environment. A big power supply company in Serbia is Jugoistok. In order to overcome the aforesaid problem there, the paper presents architecture of a platform that will collect data to be used for calculation of software quality of the information systems in Jugoistok. To overcome the problem with testing, data will be gathered in various ways – by simulation, user feedback and expert opinion.**

I. INTRODUCTION

Power supply companies are large companies responsible for managing energy usage for wide areas. That implies a large number of households that need to be served, a significant grid area, as well as highly utilized information systems (ISs) for internal business procedures, such as Customer Information Systems (CIS), Document Management System (DocMS) and Geo-Information System (GIS). Additionally, at a single power supply company level, enterprise IS are typically interconnected with various field devices, controls and metering devices within a utility-wide network. In order to enable effective monitoring and management of power supply networks according to the parameters collected (very often in real time), electric power supply companies utilize various specialized information systems, such as Supervisory Control and Data Acquisition System (SCADA), Distribution Management System (DMS) and Automatic Meter Reading (AMR), Technical Information System (TIS) [1].

An inherent property of information systems for power supply companies is about their high requirements for system quality. This includes not only hardware for such systems, but also the software and all information infrastructures (third party software, operating systems, etc.). Quality characteristics may have different definitions depending on the domain. For example, in terms of Service Oriented Architecture, they are referred as Quality of Service, which covers a wide range of techniques that match the needs of service requestors with those of the service provider's based on the network resources available [2]. In other domains quality requirements are called non-functional requirements and should be distinguished from functional requirements. The latter define what the system should do and the former put some additional conditions (in form of constraints or specifications) on how the system should perform or deliver its functionality. There are a lot of examples for quality characteristics, but the most popular are performance, reliability, usability, etc.

Monitoring and management of software quality is important for all kinds of information systems. However, at current time software quality lacks enough structured and formal support for measurement, monitoring and management [3]. The natural way for collecting data in order to determine quality of software systems is to rely on data, gathered during system testing. Nevertheless, in case of information systems for electrical supply, many of the components there are impossible to be tested in real environment, which requires other methods to be applied for such high demanding systems.

In this paper we propose a work in progress architecture for monitoring and control of software systems for power supply and electricity distribution company Jugoistok Niš (from now on Jugoistok) for one of the very important quality characteristics for such systems – namely reliability. Reliability is considered to be part of the broader notion, named dependability, which in terms of software is defined as *the ability of a computing system to deliver services that can justifiably be trusted* [2]. Besides reliability, dependability is also characterized by some other attributes, such as availability, integrity, safety, confidentiality and maintainability. The architecture we propose is service oriented, which makes it highly reusable and also applicable for other quality characteristics, not only reliability, given that one provides services, which implement the appropriate functionality.

The rest of the paper is organized as follows: Section 2 gives more information about reliability and its measurement; Section 3 makes a brief presentation of Jugoistok as a power supply company and its information system; Section 4 presents the architecture of quality management platform and finally section 5 concludes the paper and states some directions for further research.

II. SOFTWARE RELIABILITY

Generally, software reliability represents the belief we have for a system that it will not crash over a specified

period of time, given that it is operating properly at the beginning of this interval [2]. It is a statistical value and may be represented by one of the following measures:

- Probability of failure
- Failure rate
- Mean time to failure

In software engineering there exist a number of models that are generally divided in two big groups that assess software reliability. These groups are named black-box and white-box reliability models. The group of white-box models consists of several kinds of models that are used to estimate the reliability of software systems, based on the knowledge of their internal structure and processes going on inside them. On the other hand, the group of black-box models encompasses much larger number of methods that treat the software as a monolithic whole, i.e. as a black-box.

White box models are also called Architecture-Based Reliability Models (ABRMs). Usually architecture-based software reliability estimation takes the following main steps [6]:

1) Identification of computational modules (components) within software architecture;

2) Description of the actual architectural model – this includes how components are interconnected and interact with each other;

3) Definition of components failure behaviour – at this step the reliability parameters of components and their measures are identified;

4) Combination of the failure behaviour with the architectural model.

Application of white box models has a lot of advantages, among them are: ability to reuse information about reliability parameters of both the system and the components that constitute it; ability to find these modules that influence systems reliability the most, i.e; possibility to isolate and remove reliability "bottlenecks" within the system and etc. For these reasons we focus our research work on white box models.

On the other hand, black-box models take as an input some preliminary data and make statistical processing over it [4]. These models are sometimes also called reliability growth models. Reliability growth assume extensive testing of the software system and observation of failures and the time that have passed between two subsequent failures. Such data may have different representations according to the model and may be obtained by different means. When a failure is detected, the fault that caused it is removed and the process continues with the assumption that correction of the fault did not introduce new errors into the code. However, this is quite unfeasible assumption as real-world practice shows that bug fixing always introduce additional problems with the entire system and that is why regression testing is being run. As already stated in the introduction, testing all parts of the electrical supply information systems, only by testing is not always possible. Consequently, it is necessary to use other methods, not only testing for collection of input data for reliability models. Some of the other popular methods for collection of such data are: software testing, simulation, users feedback and experts opinion [5].

Simulation takes into account that it does not depend only on the structure of the software but also on the runtime information such as frequency of component reuse, execution time spent interactions between the components, etc. Users' feedback is a technique to get information about software reliability parameters of a system, by gathering data, after it has been shipped to the market and during its real usage. Data about system failures is gathered by bug reports submitted by users to a bug report subsystem and bug reports may be classified according to specific levels of severity. Experts opinion takes into account that for simple enough portions of code, reliability may be verified via code review or formal verification of source code [7].

In next section we briefly present Jugoistok as an electric supply company, and in section 4 we show the architecture, needed to monitor reliability of its information system.

III. INFORMATION SYSTEMS IN JUGOISTOK

The Jugoistok Power Supply Company in Nis, Serbia, is responsible for power management of southeast Serbia. In 2012, within the project Study on development feasibility of interoperable data exchange platform for the Jugoistok information systems, we have performed analyses of the current state of the company's information systems through existing applications, their mode of usage and internal and external communications [9]. Based on the performed analysis and its results, the importance of developing an integration solution for the company was confirmed. The main problem is that these ISs have generally not been integrated, resulting in a complex and inefficient working environment for the users. Every IS vendor in Jugoistok developed each system only to comply with requirements imposed by particular department or set of users. In order to satisfy demands from various users, each vendor tried to cover broader set of functionalities for processing large scales of different data.

A large number of business processes within electric power supply companies in Serbia, such as electric power distribution network planning, repairs, maintenance and reconfiguration, is based on the proper network model. Currently, the mentioned network models can be found in various information systems within electric power supply company. But in each information system that uses such network model, it is implemented it in a different way. Such different network modeling approaches make data exchange and manipulation between different systems within a company hard or even impossible. This produces a need for a specialized platform that should enable data exchange on electric distribution network. Usage of such platform will lower network model gap and would greatly improve the efficiency of everyday business processes related to the electric power distribution network.

Information systems in Jugoistok are mostly developed in different technologies using different platforms. Before the initial information integration has been introduced into the information system in Jugoistok, they were sharing data using point-to-point connections. Analysis has shown that in Jugoistok 20 distinct information systems are used. Majority of them (11 applications - 55%) have been developed by the IT Department of the Jugoistok while others have been developed by external partners. The same analysis has also shown that 85% of them (17

applications) use Oracle technologies, mainly Oracle database and related technologies like Oracle Forms and Oracle Reports. Besides Oracle technologies, the following technologies are also used C/C++/C#, .NET Framework, MySQL database, Perl, HTML, JavaScript, PHP, WordPress.

Analysis has also shown that majority of internal inter-information system communication is done through the shared database. This means that applications can directly access database tables, whose data are filled by other applications, or they access data through dedicated views or stored procedures. Rarely, communication is done by other, non-digital means of communication. These non-digital means of communication usually imply exchange of data written/printed on paper and then manual data insertion into the respective information system. Analysis has shown that on internal company level exist around 60 connections between applications out of which the whole 56 of them (93.33%) are connected directly through a shared database. Although it exists, percentage of manual integration and integration through dedicated communication services is negligible. All 60 identified connections between applications are point-to-point connections. This implies existence of a large number of different, specially adapted communication interfaces. Among other downsides that these point-to-point connections produce, maintaining so many dedicated communication interfaces clearly becomes a hard and error prone task.

All accounted characteristics of the information structure of the described system pose a requirement for implementation of the data integration and standard communication models.

Previously described problems produce a need for integration of various existing information systems and applications, as well as new applications, yet to be developed both inside and outside of the Jugoistok. In order to fulfill these requirements, the only solution is an implementation of information exchange infrastructure. This infrastructure needs to be adaptable and extendible so it can meet the future requirements for information integration. It needs to provide common model which can be used in different technologies and with different integration platforms.

Applications of the business information system of the Jugoistok can be organized in four logical groups, based on the functionality they carry out:

- Technical information systems
- Business information systems
- Systems that use Data Warehouse
- Web portal

Technical information systems are the systems that contain technical data on power distribution network topology, objects connected to the network and other installations. These systems are SCADA, AMR, DMS, TIS and GIS. SCADA is a system for metering, monitoring and control of transformer stations. It provides access to data in real time as well as access to archive data needed for analyses and reporting. In order to provide real time data, it generates large amounts of data about current network state. AMR is the application for remote meter reading and it provides tools for both reading and collecting data and tools for collected data validation. DMS is application for supervising, analysis, calculation and designing of the electric power distribution network. TIS is a technical information system that is used to store data on the objects and equipment of the electric power distribution network. Stored data are related to feeders, transformer stations, cables etc. GIS is a Geographical Information System for maintenance, evidencing and analysis of the electric distribution network. In Jugoistok GIS is used for mapping the whole network from the low voltage network up to high voltage network. It overlays data related to consumers, transformer stations, cables and the rest of the network infrastructure over raster maps. The whole network is shown in either schematic or line drawing modes.

Business information systems are the systems used for internal business processes of the company. They are related to employees, documenting and non-core business assets management (vehicles, buildings, HR etc.). These systems don't have anything to do with the primary business of the company.

Systems that use Data Warehouse can't be found at the moment in Jugoistok but are planned to be introduced in the near future. It will be implemented in the form of the central Data Warehouse repository, which will receive data about network and consumers from other information systems of the company. Systems that use Data Warehouse are used for advanced analyses over existing users' data and data on company's product/services usage. Based on these analyses, Data Warehouse based systems generate new information which is invaluable for future company success. Systems that will consume data from mentioned central Data Warehouse repository will be used for advanced analyses over existing consumers, their load profiles, energy consumption, electricity loses etc.

Web portal is an online presentation of the company business but also an entry point of the company's information integration. Access to data through Web portal is controlled based on different user privileges. This way, different stakeholders in and around Jugoistok Company will be able to access relevant data that can support their day to day business including company employees in various departments, regulators, external partners and others.

Previously accounted systems have become necessary in everyday Jugoistok functioning. Nevertheless, based on the changes required by different users of different systems, each of them is constantly being further developed. This constant improvement of functionalities usually makes them get deeper into the company's various business processes. Although useful, this development also produces duplication of functionalities meaning, same functionality can be found in two different systems. For instance, DMS systems, in order to properly analyze electric power distribution network need access to network technical data and therefore store them such data locally. But the same set of data, although in different format, can be found in other technical systems like TIS. Similarly, if DMS would need functionality of pinpointing the location of the network failure in order to speed up field crews dispatching, it would need to be expanded with geographic component for visualizing network elements. This geo-component would also require

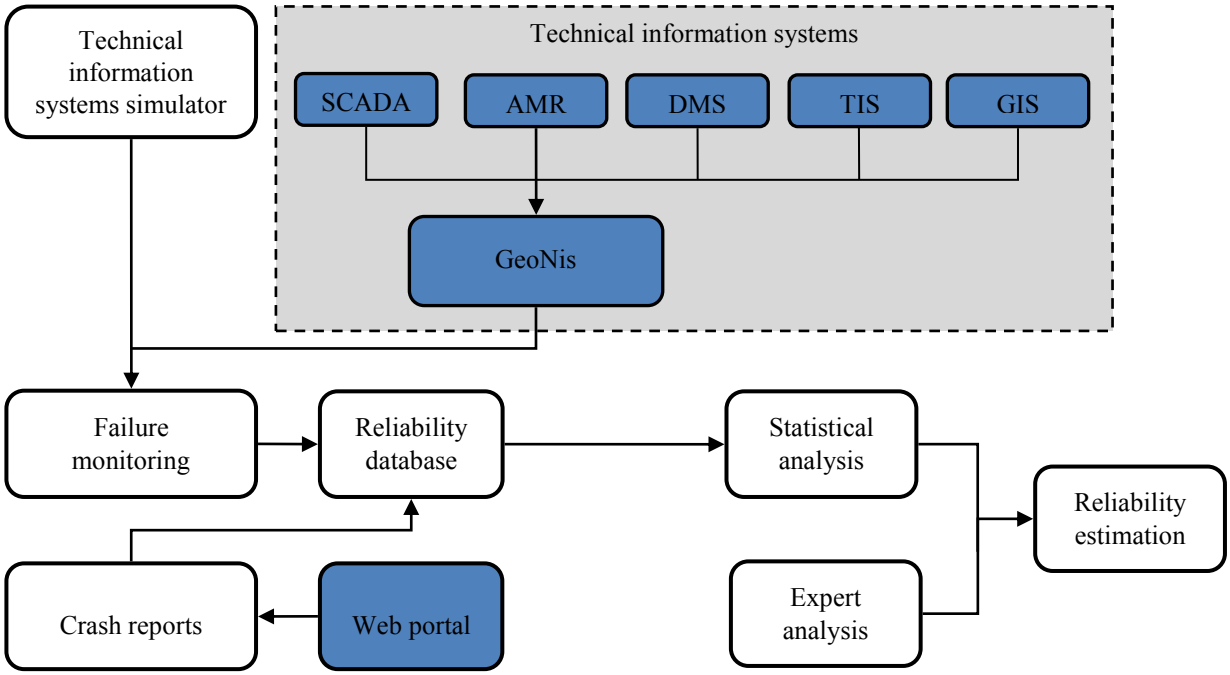


Figure 1. Architecture for software reliability management Jugoistok (Legacy software components are marked with dark background)

geographic raster maps and vector layers of the network topology as well as additional geo-analyses features which could already be found in the GIS. On the other hand, if it is required to track state and quality of the service, network events should be paired with each consumer. Source of these data are DMS and technical systems. Such data from these systems could be used, for instance, in cases when consumer requires his power to be increased.

When developing a solution for information integration in Jugoistok, we have faced a concrete requests that among other have highly prioritized data management and integration of technical subsystems with field devices, controls and metering devices (sensors). The architecture of the integration system should also include a Web Portal component, as an important integration enabler, pointing to its role and position in the communication process. The proposed solution is based on GeoNis [10], our framework for semantic interoperability. GeoNis provides information integration solution for syntax and semantic heterogeneity using hybrid ontology approach.

Information from different applications integrated using GeoNis can be published through unique Web Portal [14]. Web Portal integrates information from different IS, and displays them in a consistent, user-friendly way. In this scenario, GeoNis framework acts as buffer between the portal and other systems.

IV. ARCHITECTURE FOR QUALITY CONTROL AND MONITORING

The architecture of a system for monitoring of reliability of Jugoistok information system should have the following main components, as shown on Figure. 1:

- Technical information systems simulator – this module implements the basic workflows within the technical information systems, but the respective functionality is implemented only as stubs. Hence, if there is a mistake in the workflow it is recorded into the database via the monitoring module.
- Failure monitoring – this module gathers data about system failures during all kinds of testing of the technical information systems, including black box, integration testing, etc.
- Crash reports – this module gathers respective data from user crash reports
- Reliability database – this is the central repository module in the system. It should hold the following information: (1) Failure number; (2) Failure type (user report, simulation failure, testing failure, etc); (3) Failure severity; (4) Time elapsed after the last failure; (5) Last failure number; (6) Last failure type; (7) Particular system/component within the technical information systems, where the failure occurred. This data is used as an input for the module for statistical analysis.
- Statistical analysis – this module implements one of the black-box models for analysis of software reliability [4]. This model should be applied for each module which is was tested, simulated or user feedback has arrived for it. However, currently no model is available, which takes into account different failure types, so all failures will be regarded as one single type of failure. In that case an existing tool (for example CASRE [8]) for software reliability analysis will be wrapped as a service and used within the proposed architecture.
- Expert analysis – this module is implements a user interface for experts to input their estimates about reliability of given parts of the system which are

appropriate for such kind of reliability analysis. For example such parts are simple modules (below 300 lines of code). The expert analysis module include different forms and questionnaires that will help in formal evaluation of reliability. Development of such forms and questionnaires is part of our future research.

- Reliability estimation – this module will implement a white box model for reliability estimation, following an architectural model or business process description of the execution of the modules in technical information systems.

Figure 1 shows also the flow of information between the main modules in the architecture, as well as their connection with the respective module in existing Jugoistok information system

The presented architecture is applicable as design not only for calculation of reliability, but of wider range of quality characteristics. Indeed currently testing is a natural way to manage software quality, but as stated in the introduction, it is not always applicable to certain, critical systems. In this case the other three quality data gathering methods should be applied.

V. CONCLUSION

The paper presents a method and respective architecture to be used for evaluation of reliability of information system of electrical supply company Jugoistok in Niš, Serbia.

Main benefit of the proposed architecture is ability to monitor and predict reliability of the information system for power supply. Which in turn should be used for a number of purposes, like:

- Real time monitoring of reliability – the architecture allows implementation of specific monitoring modules, which use the output of the reliability estimation component in figure 1, to present the current state of the system, either to human operator (supervisor) or to raise an exception, which activates an alarm.
- Better service provided to users – when actively monitoring system reliability, overall number of failures will be minimized
- Increased security – if peaks in failure rate are encountered, this could be regarded as a sign of search for exploits by potential attackers to the power supply information system.

Moreover, each component in the architecture will be implemented as a service, which enforces reusability and this way, architecture may be applied for other quality characteristics, not only reliability.

Possible directions for further research include:

- Development of a model that takes into account different failure types.
- Development of appropriate questionaries' and forms for expert reliability analysis.

ACKNOWLEDGMENT

Research, presented in this paper was partially supported by the FNI 02-68/2014 project, funded by the National Science Fund, Ministry of Education and Science in Bulgaria (2014-2016).

REFERENCES

[1] Taylor, T.; Kazemzadeh, H. Integrated SCADA/DMS/OMS: Increasing distribution operations efficiency. *Electr. Energy T&D Mag.* 2009, 9, 32–34.

[2] Avižienis, A., Laprie, J-C., Randell, B.: Basic concepts and Taxonomy of dependable and secure computing, *IEEE Trans on Dependable and Secure computing*, Vol. 1, Issue 1, Jan -March 2004.

[3] Dimov A. Measurement of Software Quality Characteristics. *Computer Science and Technologies*. Bulgaria, Vol. 1, 2013. 199-204.

[4] Farr, W., Software reliability modeling survey, in: M.R. Lyu (Ed.), *Handbook of Software Reliability Engineering*, McGraw-Hill, New York, 1996, pp. 71–117.

[5] Dimov, A., S. Chandran and S. Punnekkat. (2010). How do we Collect Data for Software Reliability Estimation? Proceedings of the *11th International Conference on Computer Systems and Technologies* (CompSysTech). ACM ICPS, vol. 471. Sofia, Bulgaria. June 17-18, 2010.

[6] Gokhale, S., Architecture-Based Software Reliability Analysis: Overview and Limitations, In *IEEE Transactions on Dependable Security Computing* 4(1): 32-40 (2007).

[7] Arun Babu, P., C. Senthil Kumar, and N. Murali. "A hybrid approach to quantify software reliability in nuclear safety systems." *Annals of Nuclear Energy* 50 (2012): 133-140.

[8] Nikora, A., Computer Aided Software Reliability Estimation User's Guide (CASRE), Ver-sion 3.0, 2002.

[9] Stoimenov, L. et al, M. *Study on Development Feasibility of Interoperable Data Exchange Platform for the Information System ED Jugoistok*;Technical report, Niš, Serbia, October 2012

[10] L. Stoimenov, S. Đorđević-Kajan, "An Architecture for Interoperable GIS Use in a Local Community Environment", Computers & Geosicence, Elsevier, 2005, Vol. 31, No. 2, pp.211-220, March 2005

Model Integration for Territorial Environmental & Social Assessment trough Life-Cycle Approach: The case study of the Province of Matera.

Francesca Intini *, Nicola Cardinale *, Michele Dassisti **, Alexis Aubry, Hervé Panetto (***)
* DICEM, Università degli Studi della Basilicata, Italy - {francesca.intini; nicola.cardinale}@unibas.it
** DMMM, Politecnico di Bari, Viale Japigia 182, 70126 - Bari, Italy - michele.dassisti@poliba.it
*** Centre de Recherche en Automatique de Nancy Université de Lorraine - {alexis.aubry; herve.panetto}@univ-lorraine.fr

Abstract—**Systemic view in Strategic Environmental Management is gaining even more attention. The paper proposes the case of a Territorial Environmental and Social Assessment (TESA), based on a Life-cycle approach, applied to set-up the Matera Provincial Plan for a strategic environmental management. Information for performing LCA is here integrated through an Energy-Social Planning model as a decisional support tool. The city-owned buildings are critical facilities for Territorial energy management: the focus here is mainly on energy efficiency. The results obtained of the study contribute to reach EU targets according to the new Directive of the European Parliament and of the Council on the energy performance of buildings.**

I. INTRODUCTION

A Strategic Environmental Assessment (SEA) is an approach to incorporate environmental issues in territorial plan and program development. It can also be regarded as a decision support process, especially when considering plan development. Details on SEA are defined in the Directive 2001/42/EC, which is mandatory for the EU Member States by July 2004.

Different analytical tools can be used to perform such assessments (e.g., Material Flow Analysis, the Ecological Footprint, Emergy analysis, etc.). Among all, Life Cycle Assessment (LCA) has been identified as one of the most promising tools, as it can be used to perform a comprehensive assessment of a territory as a whole (systemic view). Behind LCA, there is the assessment of the environmental impacts and resources used throughout a system's life cycle; for a product, for instance, from raw material acquisition, toward production and use phases, up to waste management. Its capability to avoid problem-shifting between life cycle stages, territories, and environmental impacts is a significant asset [1].

Few studies has been devoted so far performing LCA at a territorial scale, to assess the impacts of specific anthropic activities (economy, social, cultural, training, education,..). An approach based on LCA has been developed to provide macro-level life-cycle indicators to monitor the consumption of the EU-27 and Germany (European Commission 2012a). LCA study has been performed to evaluate different energy resource-management scenarios as in Sweden municipalities [2]; the same has been done for water systems in Sydney [3]. Another study investigate the environmental sustainability of the Province of Siena and of its communes, by means of different indicators and methods of analysis [5]. The standardized LCA framework has never been applied as such to study a territorial system [4].

In this paper, an integrated approach of a Life-Cycle Assessment methodology is proposed to perform a Territorial Environmental & Social Assessment (TESA) to provide a systemic analysis to the Matera strategic environmental management based on Life-cycle policy. A conceptual model is proposed to contextualize the data and assessment criteria and make the TESA congruent with the on-going Territorial energy strategy. This allowed to take into account the meaning of data in a dynamic scenario, thus allowing to set-up an effective analytical tool. A benchmark is also provided with a previous Social & Environmental Management applied to the municipal energy-plan of the Province of Matera.

When estimating environmental impacts of a territory, one can consider it as a "black box" that interacts with other black-box territories via a variety of inputs and outputs. In this case, impacts of human activities are independent of location within the territory. The territory should be considered as a system in which emissions occur at different places and impacts are influenced by the sensitivity of the receiving environment [6].

II. LCA AND TESA PROCESS

This section outlines the design and scope of the Territorial Environmental & Social Assessment (TESA) process performed according to the LCA approach made possible by the use of a conceptual Model as shown in figure 1.

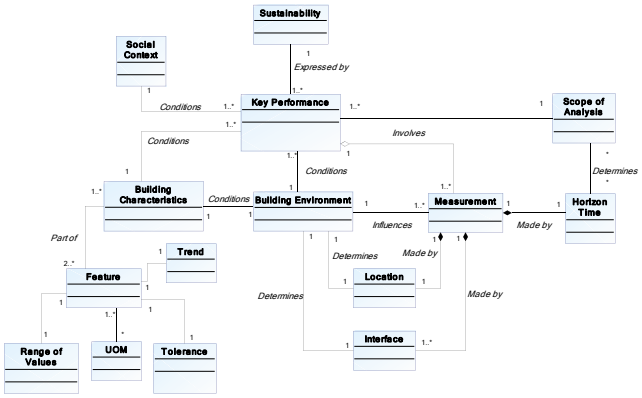


Figure 1. TESA Conceptual Model

It is clear that the core of the approach, the assessment of sustainability, is based on the appropriate selection of one or more key performance indicators. In the same way, it is evident that these are influenced by some factors (namely the scope of the analysis and the social context) which should be taken into account. Buildings features are critical to performance assessment; this fact should be also considered when organizing data or information for the assessment. To say it in a word, the conceptual model represents a sort of guide to organize the overall information gathering and organization at Territorial level. This process is still under testing and the present paper represents an initial method to go in the direction of organizing a systemic approach to the area.

Where sustainability is the core concept to the successful assessment. It is more than an index since it is expressed by a set of key performance indicators. Building environment and building feature as well as social context strongly influences the assessment approch. The same model specifies the critical knowledge assets either to perform a correct Territorial Environmental & Social Assessment as well as to design and manage an appropriate and interoperable database at territorial level. The model in fact ensures the internal congruence of information and knowledge of several Territorial real estate assets so as to build a de-facto interoperable management database.

The process consists of several steps; the focus will be on the city-owned buildings are an important asset for managing the Environmental plan of a Region, where it is necessary to strongly focus on energy efficiency.

In the following we give details of the case study presented to discuss model integration.

A. First TESA- process step: Energy analysis of public buildings

The first phase of the TESA process based on the Territorial LCA consists of a careful check-up (inventory) of the municipal construction, to provide a detailed picture of the public buildings features, their facilities as well as their energy performances. This phase is a key element for maintenance planning aimed at energy saving, both on the buildings as well as the energy plants. This Territorial inventory of public buildings is strongly influenced by the Territorial management strategies in terms of Social welfare as well as Energy management. Strategic choices in this terms strongly influence the data related to energy consumption of different buildings (say, efficiency improvement plans, energy policies in term of renewable sources, etc.). TESA may allow to identify the critical in territorial management and therefore may enable to support the development of a intervention strategy on the entire building stock and energy plant, selecting the priority actions to take.

Energy analysis of public buildings was performed to get the relevant data to the energy performance of the Territorial public buildings as well as the Territorial energy facilities (thermal characteristics of the housing, the energy conversion systems and distribution etc.). Consumption (bills), characteristics structural building, near buildings (away) as well as areas of shading were also evaluated. The Energy-Social Planning model was adopted to guide the data acquisition based on the information related to each asset: namely social destination of buildings (and therefore the consequent use) as well as the forecasted improvements on the buildings themselves. The social destination are: schools, municipalities, hospitals, theaters, stadiums,..

B. Second TESA- process step: Data interoperability using the TESA Model

The second phase of SEA process includes the performance of energy-audit of municipal buildings to assess the potential energy savings on municipal assets. The energy-audit survey (energy) involves an on-field inspection of the buildings and the collection of detailed information on the energy efficiency of the building envelope, as well as their facilities and equipments for energy production. Following the audit it is easy to assess savings and possible maintenances based on decision-support tools (say, e.g. cost-benefit analysis).

The basic variables that are behind the TESA Model and that are important to be cataloged for buildings and interact have been devised as follows:

- GPS coordinates;
- Consumption of methane gas;
- Consumption of thermal energy;
- Type of heating system;
- Heat fuel value
- heating system efficiency;
- Year of construction;
- Wall Building Materials (brick / concrete / stone / mixed)
- historical-artistic constraints (yes / no)
- Number of floors (including ground floor)
- Building height (m)
- Usable living areas (square meter)
- Heated volume (cubic meter)
- Type of roofs /coverages (plane / pitched)
- Windowed area (square meter)
- Matte surfaces (square meter)
- No. of lightings
- Renewable energy production
- Neighborhood shading

The consumption data are recorded on annual basis and were referred to residents of a given property, identified with the cadastral data according to the social planning information. These data have a format incongruence: it was necessary to link them graphically to the Geographic Information System with the aim of having a map view of the territory and recovering these using data of real estate as search keys. These interoperability problems were solved by merging three different databases (Land Registry property, civil status, energy consumption) according to the Energy-Social Model as reference. This allowed to create a common basis to have a wide systemic view of the territory to monitor the energy consumption.

C. Third TESA process step: Queries for LCA data

The third phase of the SEA process here performed focused on municipal assets involves the identification and assessment of possible maintenance or improvement interventions on the buildings. The interventions can be identified according to an integrated energy approach, which includes: measures of thermal insulation for buildings, the application of advanced technologies for shading, ventilation, heat recovery, heating and summer cooling high efficiency, possible use of renewable fuels such as wood or vegetable oil and the use of solar energy active and passive, and finally, the adoption of electrical equipments with low consumption. The various measures can be combined to assess the most promising cost-effective mix.

Starting from the common database created in step 2 several queries were possible to identify nearby buildings to endeavor their interaction with integrated operations to generate economic and social benefits in a systemic perspective.

The selection of appropriate technologies and solutions for energy savings can then be performed after the LCA study according to the environmental parameters devised. In this case the LCA integrated with energy, economic and social variables determines the optimal solution, to determine the possible operations of the system, avoiding wastage.

III. THE TERRITORIAL PLANNING

The Province of Matera (Italy) signed on 2010, the Partnership Agreement with the Directorate General of Energy Commission, assuming the role of "Territorial Coordinator" and “Supporting Structure” of the EC for the territory of the Matera Province, undertaking to:

- promote the adherence to the Covenant of Mayors of the municipalities of the territory and provide support and coordination to the municipalities who already have signed the Covenant of Mayors;
- provide technical and strategic assistance to the municipalities that wish to join the Covenant of Mayors but which do not have the resources to prepare an action plan for sustainable energy;
- provide financial support to municipalities and opportunities for the development and the implementation of the Sustainable Energy Action Plan.

The main outcomes of the present SEA process was to coordinate the Energy Efficiency and Renewable Energy switching interventions to reach an unique investment portfolio to provide important savings in Energy Territorial government. According to the assessment results derived from the SEA process, the investment portfolio can be composed in a near future of a sum of smaller interventions (building groups) each composed of two/three buildings. The test case has been Matera school complexes according to a widespread international paradigm of energy districts or settlements. Endeavoring a mix of technological solutions, it will be possible to optimize the interaction between local energy generations and consumptions, reducing energy consumption and using as much as possible and economically compatible renewable sources in the LCA view. The integrated view of public buildings descending from the SEA approach proposed will allow to optimize the design of the entire public real-estate system by acting simultaneously on the minimization of the consumption of individual households, on local and economical production of energy, the integration of renewable energy sources and the efficient management of the system. The proposed SEA process may suggest more typologies of interventions on different kind of building complexes: say, for instance, between school buildings with public offices or school buildings with hospitals.

These examples will represent case examples (or pilots) to define the guidelines of virtuous actions that can be performed for energy management in the Matera province. The aim is also to promote other interventions in a smart grids logic (in line with the guidelines laid down in this area by the “National strategy for the internal areas”) either among different public authorities or between public and private actors (taking into account the specific utilization of schools energy facilities that are used above all in the morning and from September to May).

Following to the SEA process performed, an archive with all the structural information of the buildings within the jurisdiction of the Province of Matera was drawn up. Data coming from a wide auditing project were also performed within the same province, taking diagnoses, testing and certificating the energy and environmental buildings. This corpus of knowledge was used for planning of the energy retrofit of the proposed buildings.

A. Geographical Area

The geographical zone in which the proposed action will be implemented concerns the City of Matera located in the convergence region of Basilicata, in southern Italy.

Matera is the capital of the province of Matera and has a territory of about 387,4 sqkm with a population of 60.023 inhabitants (the density is 150 inhabitants/sqkm).

Regarding the territory involved, the challenge of the prosed action is to actively involve all local and Territorial stakeholders and to achieve a significant impact in terms of contribution to the realization of the Cohesion Policy of the European Commission.

For the present study no attention will be devoted to the social aspect because they aren’t the focus of this paper.

In the context of the proposed action it is important to distinguish between two important intervention areas: “cities” and “internal areas”. This distinction particularly represents the Italian reality. The largest part of the Italian

territory is in fact characterized by a spatial organization based on small centers, usually of reduced geographical dimensions and capable to ensure to its inhabitants only a limited accessibility to essential services. The action proposed aims to actively contribute directly at one of the five development factors – energy conservation and renewable energy – by representing a "pilot" action with high replication capacity which will contribute to increase the Territorial cohesion and access to internal market, maintaining at the same time the "specificity" of such areas.

B. Analysed buildings

The following description provides an overview of the interested buildings, grouped in complexes of three buildings.

The school complex, located in Matera, includes:

- the Liceo Scientifico "Dante Alighieri" (LS);
- the Liceo Classico "Emanuele Duni" (LC);
- the Istituto Tecnico Commerciale "Loperfido-Olivetti" (ITC2)

The LS is characterized by numerous classrooms for teaching activity with large windows. It has 10 laboratories, 1600 sqm large gym, a library and an assembly hall. In the school year 2013-2014 there was about 793 pupils divided into 32 classes.

The LC of Matera is the oldest school of the city (established in 1864). The current headquarters in the Nazioni Unite Street has 4 laboratories, 2 gyms and a library. In the school year 2013-2014 there was 524 pupils divided into 22 classes.

With regard to the ITC2, the building located in Aldo Moro Street will be involved in the energy intervention of the second school complex.

The figure 2 below shows the closed geographical location of the three school complexes.

FIGURE 2. Geographical location of the second school complex

The LS was built in 1971 in reinforced concrete, four-floors with a flat roof; the building height is of 12 m, with a surface area of 7.830 m2 and a heated volume of 27.400 cubic meters (m3).

The LC was built in 1966 in reinforced concrete, six-floors with a pitched roof, the building height is of 20 m, with a surface area of 6.085 m2 and a heated volume of 21.800 m3.

The ITC2 in Aldo Moro Street was built in 1961 in reinforced concrete, five-floors with a flat roof, the building height is of 18 m, with a surface area of 5.814 m2 and a heated volume of 18.600 m3.

The LS is equipped with a heating system fueled with natural gas boiler, whose annual consumption in 2012 amounted to 57.680 m3. The school has not insulated walls and ceilings; the glazed surfaces, of approximately 956 m2, are single-glazed windows, with iron frame without thermal break. In 2012 the electricity consumption were equal to 79.242 kWh, whose main component is given by lighting with about 439 neon lighting.

The LC is equipped with a heating system fueled with natural gas boiler, whose annual consumption in 2012 amounted to 23.546 m3. The institute has insulated walls and ceilings according to the Territorial energy plan but the glazed surfaces of approximately 155 m2 are represented by single-glazed windows, with iron frame without thermal break. In 2012 the electricity consumption totaled 39.532 kWh, whose main component is given by lighting with about 366 neon lighting.

The ITC2 is equipped with a heating system fueled with natural gas boiler, whose annual consumption in 2012 amounted to 33.193 m3. The institute does not have insulated walls and ceilings; the glazed surfaces of approximately 777 m2 are represented by single-glazed windows, with iron frame without thermal break. In 2012 the electricity consumption totaled 75.579 kWh, whose main component is given by lighting with about 384 neon lighting.

Energy-Social Planning serves to optimize investments in the territory increasing social activity and energy.

For these three important school complexes several LCA-optimal energy improvements have been devised according to the Energy-Social Planning following the SEA process:

- thermal coating of the opaque part of the building envelope in order to reduce the loss of heat with the external environment and increase the thermal insulation, only for the ITC2 and LC;
- replacement of windows with low-emissivity double-glazed windows;
- installation of photovoltaic solar shading on the south facades of the buildings allowing to make the best use of free solar gains and try to eliminate or minimize them when they can be harmful, avoiding overheating in summer and ensuring the efficient use of natural lighting during the winter. It is estimated a production of 12 kW for 100 m2 of solar shading.
- replacement of obsolete lighting equipment with LED type with built-in sensors for monitoring the presence and adjustment in function of the natural light, ensuring this way at least a 60% of savings, with the aim of reducing energy consumption and costs;
- replacement of the existing boiler with a new next-generation condensing boiler, with efficiency ratios exceeding 107% suitable for

operation with gas burners and installation of thermostatic valves where necessary.

The investment required for these projects amount to approximately €1.131.000,00. The interventions generates an annual saving of 100% of electricity and save 56% of natural gas. It is also reasonable to fed into the grid part of RES energy produced.

TABLE1.
DETAILED DESCRIPTION OF THE PROPOSED INVESTMENTS

Location and name of buildings	**Province of Matera LS, LC, ITC2**
Number of buildings	3
Total surface (m2)	19.729
Current primary electricity energy consumption (MWh/year)	431
Current natural gas consumption(MWh/year)	1.211
Primary electricity energy savings (MWh/year)	872
Natural gas savings (MWh/year)	676
Energy savings %	100%
Natural gas savings %	56%
Average GHG emissions (tCO2e/m2/year)	321,27
Estimated CO2 reduction %	90%

IV. CONCLUSIONS

The proposed approach of TESA process based on LCA applied to a real case of school complexes, led to an interesting paradigm of energy districts in which, through a mix of technological solutions, Energy Efficiency and RES interventions can be rejoined into an unique investment. This allows to optimize the interaction between local energy generation and consumption.

REFERENCES

[1] G. Finnveden, MZ Hauschild, T Ekvall, J Guinee, R. Heijungs, S. Hellweg, A. Koehler, D. Pennington, S. Suh, "Recent Developments in Life Cycle Assessment", *J Environ Manage*, 91 (1): 1–21, 2009

[2] A. Björklund, "Life cycle assessment as an analytical tool in strategic environmental assessment. Lessons learned from a case study on municipal energy planning in Sweden", *Environ Impact Assess Rev* 32:82–87, 2012

[3] S. Lundie, GM Peters, PC Beavis, "Life cycle assessment for sustainable metropolitan water systems planning", *Environ Sci Technol* 38:3465–3473, 2004.

[4] E. Loiseau, G. Junqua, P. Roux, V. Bellon-Maurel, "Environmental assessment of a territory: an overview of existing tools and methods", *J Environ Manage* 112:213–225, 2012.

[5] M. Bagliania, A. Galli. V. Niccolucci, N. Marchettini, "Ecological footprint analysis applied to a sub-national area: The case of the Province of Siena (Italy)", *Journal of Environmental Management* 86,354–364, 2008.

[6] Nitschelm, L. et al. ,"Utility of spatially explicit LCA for agricultural territories", *Proceedings of the 9th International Conference on Life Cycle Assessment in the Agri-Food Sector*,2014.

Model Integration for Territorial Environmental & Social Assessment trough Life-Cycle Approach: The case study of the Province of Matera.

Francesca Intini *, Nicola Cardinale *, Michele Dassisti **, Alexis Aubry, Hervé Panetto (***)
* DICEM, Università degli Studi della Basilicata, Italy - {francesca.intini; nicola.cardinale}@unibas.it
** DMMM, Politecnico di Bari, Viale Japigia 182, 70126 - Bari, Italy - michele.dassisti@poliba.it
*** Centre de Recherche en Automatique de Nancy Université de Lorraine - {alexis.aubry; herve.panetto}@univ-lorraine.fr

Abstract—**Systemic view in Strategic Environmental Management is gaining even more attention. The paper proposes the case of a Territorial Environmental and Social Assessment (TESA), based on a Life-cycle approach, applied to set-up the Matera Provincial Plan for a strategic environmental management. Information for performing LCA is here integrated through an Energy-Social Planning model as a decisional support tool. The city-owned buildings are critical facilities for Territorial energy management: the focus here is mainly on energy efficiency. The results obtained of the study contribute to reach EU targets according to the new Directive of the European Parliament and of the Council on the energy performance of buildings.**

I. INTRODUCTION

A Strategic Environmental Assessment (SEA) is an approach to incorporate environmental issues in territorial plan and program development. It can also be regarded as a decision support process, especially when considering plan development. Details on SEA are defined in the Directive 2001/42/EC, which is mandatory for the EU Member States by July 2004.

Different analytical tools can be used to perform such assessments (e.g., Material Flow Analysis, the Ecological Footprint, Emergy analysis, etc.). Among all, Life Cycle Assessment (LCA) has been identified as one of the most promising tools, as it can be used to perform a comprehensive assessment of a territory as a whole (systemic view). Behind LCA, there is the assessment of the environmental impacts and resources used throughout a system's life cycle; for a product, for instance, from raw material acquisition, toward production and use phases, up to waste management. Its capability to avoid problem-shifting between life cycle stages, territories, and environmental impacts is a significant asset [1].

Few studies has been devoted so far performing LCA at a territorial scale, to assess the impacts of specific anthropic activities (economy, social, cultural, training, education,..). An approach based on LCA has been developed to provide macro-level life-cycle indicators to monitor the consumption of the EU-27 and Germany (European Commission 2012a). LCA study has been performed to evaluate different energy resource-management scenarios as in Sweden municipalities [2]; the same has been done for water systems in Sydney [3]. Another study investigate the environmental sustainability of the Province of Siena and of its communes, by means of different indicators and methods of analysis [5]. The standardized LCA framework has never been applied as such to study a territorial system [4].

In this paper, an integrated approach of a Life-Cycle Assessment methodology is proposed to perform a Territorial Environmental & Social Assessment (TESA) to provide a systemic analysis to the Matera strategic environmental management based on Life-cycle policy. A conceptual model is proposed to contextualize the data and assessment criteria and make the TESA congruent with the on-going Territorial energy strategy. This allowed to take into account the meaning of data in a dynamic scenario, thus allowing to set-up an effective analytical tool. A benchmark is also provided with a previous Social & Environmental Management applied to the municipal energy-plan of the Province of Matera.

When estimating environmental impacts of a territory, one can consider it as a "black box" that interacts with other black-box territories via a variety of inputs and outputs. In this case, impacts of human activities are independent of location within the territory. The territory should be considered as a system in which emissions occur at different places and impacts are influenced by the sensitivity of the receiving environment [6].

II. LCA AND TESA PROCESS

This section outlines the design and scope of the Territorial Environmental & Social Assessment (TESA) process performed according to the LCA approach made possible by the use of a conceptual Model as shown in figure 1.

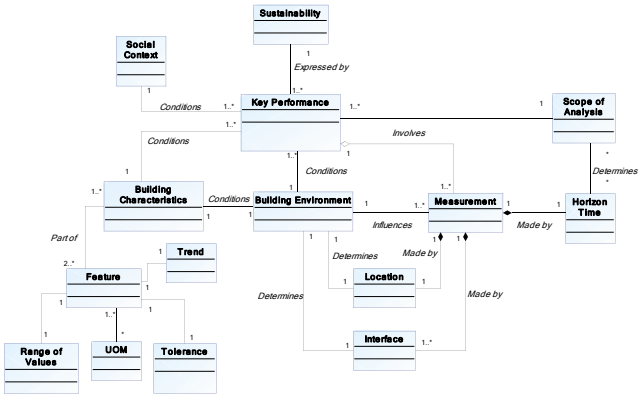


Figure 1. TESA Conceptual Model

It is clear that the core of the approach, the assessment of sustainability, is based on the appropriate selection of one or more key performance indicators. In the same way, it is evident that these are influenced by some factors (namely the scope of the analysis and the social context) which should be taken into account. Buildings features are critical to performance assessment; this fact should be also considered when organizing data or information for the assessment. To say it in a word, the conceptual model represents a sort of guide to organize the overall information gathering and organization at Territorial level. This process is still under testing and the present paper represents an initial method to go in the direction of organizing a systemic approach to the area.

Where sustainability is the core concept to the successful assessment. It is more than an index since it is expressed by a set of key performance indicators. Building environment and building feature as well as social context strongly influences the assessment approch. The same model specifies the critical knowledge assets either to perform a correct Territorial Environmental & Social Assessment as well as to design and manage an appropriate and interoperable database at territorial level. The model in fact ensures the internal congruence of information and knowledge of several Territorial real estate assets so as to build a de-facto interoperable management database.

The process consists of several steps; the focus will be on the city-owned buildings are an important asset for managing the Environmental plan of a Region, where it is necessary to strongly focus on energy efficiency.

In the following we give details of the case study presented to discuss model integration.

A. First TESA- process step: Energy analysis of public buildings

The first phase of the TESA process based on the Territorial LCA consists of a careful check-up (inventory) of the municipal construction, to provide a detailed picture of the public buildings features, their facilities as well as their energy performances. This phase is a key element for maintenance planning aimed at energy saving, both on the buildings as well as the energy plants. This Territorial inventory of public buildings is strongly influenced by the Territorial management strategies in terms of Social welfare as well as Energy management. Strategic choices in this terms strongly influence the data related to energy consumption of different buildings (say, efficiency improvement plans, energy policies in term of renewable sources, etc.). TESA may allow to identify the critical in territorial management and therefore may enable to support the development of a intervention strategy on the entire building stock and energy plant, selecting the priority actions to take.

Energy analysis of public buildings was performed to get the relevant data to the energy performance of the Territorial public buildings as well as the Territorial energy facilities (thermal characteristics of the housing, the energy conversion systems and distribution etc.). Consumption (bills), characteristics structural building, near buildings (away) as well as areas of shading were also evaluated. The Energy-Social Planning model was adopted to guide the data acquisition based on the information related to each asset: namely social destination of buildings (and therefore the consequent use) as well as the forecasted improvements on the buildings themselves. The social destination are: schools, municipalities, hospitals, theaters, stadiums,..

B. Second TESA- process step: Data interoperability using the TESA Model

The second phase of SEA process includes the performance of energy-audit of municipal buildings to assess the potential energy savings on municipal assets. The energy-audit survey (energy) involves an on-field inspection of the buildings and the collection of detailed information on the energy efficiency of the building envelope, as well as their facilities and equipments for energy production. Following the audit it is easy to assess savings and possible maintenances based on decision-support tools (say, e.g. cost-benefit analysis).

The basic variables that are behind the TESA Model and that are important to be cataloged for buildings and interact have been devised as follows:

- GPS coordinates;
- Consumption of methane gas;
- Consumption of thermal energy;
- Type of heating system;
- Heat fuel value
- heating system efficiency;
- Year of construction;
- Wall Building Materials (brick / concrete / stone / mixed)
- historical-artistic constraints (yes / no)
- Number of floors (including ground floor)
- Building height (m)
- Usable living areas (square meter)
- Heated volume (cubic meter)
- Type of roofs /coverages (plane / pitched)
- Windowed area (square meter)
- Matte surfaces (square meter)
- No. of lightings
- Renewable energy production
- Neighborhood shading

The consumption data are recorded on annual basis and were referred to residents of a given property, identified with the cadastral data according to the social planning information. These data have a format incongruence: it was necessary to link them graphically to the Geographic Information System with the aim of having a map view of the territory and recovering these using data of real estate as search keys. These interoperability problems were solved by merging three different databases (Land Registry property, civil status, energy consumption) according to the Energy-Social Model as reference. This allowed to create a common basis to have a wide systemic view of the territory to monitor the energy consumption.

C. Third TESA process step: Queries for LCA data

The third phase of the SEA process here performed focused on municipal assets involves the identification and assessment of possible maintenance or improvement interventions on the buildings. The interventions can be identified according to an integrated energy approach, which includes: measures of thermal insulation for buildings, the application of advanced technologies for shading, ventilation, heat recovery, heating and summer cooling high efficiency, possible use of renewable fuels such as wood or vegetable oil and the use of solar energy active and passive, and finally, the adoption of electrical equipments with low consumption. The various measures can be combined to assess the most promising cost-effective mix.

Starting from the common database created in step 2 several queries were possible to identify nearby buildings to endeavor their interaction with integrated operations to generate economic and social benefits in a systemic perspective.

The selection of appropriate technologies and solutions for energy savings can then be performed after the LCA study according to the environmental parameters devised. In this case the LCA integrated with energy, economic and social variables determines the optimal solution, to determine the possible operations of the system, avoiding wastage.

III. THE TERRITORIAL PLANNING

The Province of Matera (Italy) signed on 2010, the Partnership Agreement with the Directorate General of Energy Commission, assuming the role of "Territorial Coordinator" and “Supporting Structure” of the EC for the territory of the Matera Province, undertaking to:

- promote the adherence to the Covenant of Mayors of the municipalities of the territory and provide support and coordination to the municipalities who already have signed the Covenant of Mayors;
- provide technical and strategic assistance to the municipalities that wish to join the Covenant of Mayors but which do not have the resources to prepare an action plan for sustainable energy;
- provide financial support to municipalities and opportunities for the development and the implementation of the Sustainable Energy Action Plan.

The main outcomes of the present SEA process was to coordinate the Energy Efficiency and Renewable Energy switching interventions to reach an unique investment portfolio to provide important savings in Energy Territorial government. According to the assessment results derived from the SEA process, the investment portfolio can be composed in a near future of a sum of smaller interventions (building groups) each composed of two/three buildings. The test case has been Matera school complexes according to a widespread international paradigm of energy districts or settlements. Endeavoring a mix of technological solutions, it will be possible to optimize the interaction between local energy generations and consumptions, reducing energy consumption and using as much as possible and economically compatible renewable sources in the LCA view. The integrated view of public buildings descending from the SEA approach proposed will allow to optimize the design of the entire public real-estate system by acting simultaneously on the minimization of the consumption of individual households, on local and economical production of energy, the integration of renewable energy sources and the efficient management of the system. The proposed SEA process may suggest more typologies of interventions on different kind of building complexes: say, for instance, between school buildings with public offices or school buildings with hospitals.

These examples will represent case examples (or pilots) to define the guidelines of virtuous actions that can be performed for energy management in the Matera province. The aim is also to promote other interventions in a smart grids logic (in line with the guidelines laid down in this area by the “National strategy for the internal areas”) either among different public authorities or between public and private actors (taking into account the specific utilization of schools energy facilities that are used above all in the morning and from September to May).

Following to the SEA process performed, an archive with all the structural information of the buildings within the jurisdiction of the Province of Matera was drawn up. Data coming from a wide auditing project were also performed within the same province, taking diagnoses, testing and certificating the energy and environmental buildings. This corpus of knowledge was used for planning of the energy retrofit of the proposed buildings.

A. Geographical Area

The geographical zone in which the proposed action will be implemented concerns the City of Matera located in the convergence region of Basilicata, in southern Italy.

Matera is the capital of the province of Matera and has a territory of about 387,4 sqkm with a population of 60.023 inhabitants (the density is 150 inhabitants/sqkm).

Regarding the territory involved, the challenge of the prosed action is to actively involve all local and Territorial stakeholders and to achieve a significant impact in terms of contribution to the realization of the Cohesion Policy of the European Commission.

For the present study no attention will be devoted to the social aspect because they aren’t the focus of this paper.

In the context of the proposed action it is important to distinguish between two important intervention areas: “cities” and “internal areas”. This distinction particularly represents the Italian reality. The largest part of the Italian

territory is in fact characterized by a spatial organization based on small centers, usually of reduced geographical dimensions and capable to ensure to its inhabitants only a limited accessibility to essential services. The action proposed aims to actively contribute directly at one of the five development factors – energy conservation and renewable energy – by representing a "pilot" action with high replication capacity which will contribute to increase the Territorial cohesion and access to internal market, maintaining at the same time the "specificity" of such areas.

B. Analysed buildings

The following description provides an overview of the interested buildings, grouped in complexes of three buildings.

The school complex, located in Matera, includes:

- the Liceo Scientifico "Dante Alighieri" (LS);
- the Liceo Classico "Emanuele Duni" (LC);
- the Istituto Tecnico Commerciale "Loperfido-Olivetti" (ITC2)

The LS is characterized by numerous classrooms for teaching activity with large windows. It has 10 laboratories, 1600 sqm large gym, a library and an assembly hall. In the school year 2013-2014 there was about 793 pupils divided into 32 classes.

The LC of Matera is the oldest school of the city (established in 1864). The current headquarters in the Nazioni Unite Street has 4 laboratories, 2 gyms and a library. In the school year 2013-2014 there was 524 pupils divided into 22 classes.

With regard to the ITC2, the building located in Aldo Moro Street will be involved in the energy intervention of the second school complex.

The figure 2 below shows the closed geographical location of the three school complexes.

FIGURE 2. Geographical location of the second school complex

The LS was built in 1971 in reinforced concrete, four-floors with a flat roof; the building height is of 12 m, with a surface area of 7.830 m2 and a heated volume of 27.400 cubic meters (m3).

The LC was built in 1966 in reinforced concrete, six-floors with a pitched roof, the building height is of 20 m, with a surface area of 6.085 m2 and a heated volume of 21.800 m3.

The ITC2 in Aldo Moro Street was built in 1961 in reinforced concrete, five-floors with a flat roof, the building height is of 18 m, with a surface area of 5.814 m2 and a heated volume of 18.600 m3.

The LS is equipped with a heating system fueled with natural gas boiler, whose annual consumption in 2012 amounted to 57.680 m3. The school has not insulated walls and ceilings; the glazed surfaces, of approximately 956 m2, are single-glazed windows, with iron frame without thermal break. In 2012 the electricity consumption were equal to 79.242 kWh, whose main component is given by lighting with about 439 neon lighting.

The LC is equipped with a heating system fueled with natural gas boiler, whose annual consumption in 2012 amounted to 23.546 m3. The institute has insulated walls and ceilings according to the Territorial energy plan but the glazed surfaces of approximately 155 m2 are represented by single-glazed windows, with iron frame without thermal break. In 2012 the electricity consumption totaled 39.532 kWh, whose main component is given by lighting with about 366 neon lighting.

The ITC2 is equipped with a heating system fueled with natural gas boiler, whose annual consumption in 2012 amounted to 33.193 m3. The institute does not have insulated walls and ceilings; the glazed surfaces of approximately 777 m2 are represented by single-glazed windows, with iron frame without thermal break. In 2012 the electricity consumption totaled 75.579 kWh, whose main component is given by lighting with about 384 neon lighting.

Energy-Social Planning serves to optimize investments in the territory increasing social activity and energy.

For these three important school complexes several LCA-optimal energy improvements have been devised according to the Energy-Social Planning following the SEA process:

- thermal coating of the opaque part of the building envelope in order to reduce the loss of heat with the external environment and increase the thermal insulation, only for the ITC2 and LC;
- replacement of windows with low-emissivity double-glazed windows;
- installation of photovoltaic solar shading on the south facades of the buildings allowing to make the best use of free solar gains and try to eliminate or minimize them when they can be harmful, avoiding overheating in summer and ensuring the efficient use of natural lighting during the winter. It is estimated a production of 12 kW for 100 m2 of solar shading.
- replacement of obsolete lighting equipment with LED type with built-in sensors for monitoring the presence and adjustment in function of the natural light, ensuring this way at least a 60% of savings, with the aim of reducing energy consumption and costs;
- replacement of the existing boiler with a new next-generation condensing boiler, with efficiency ratios exceeding 107% suitable for

operation with gas burners and installation of thermostatic valves where necessary.

The investment required for these projects amount to approximately €1.131.000,00. The interventions generates an annual saving of 100% of electricity and save 56% of natural gas. It is also reasonable to fed into the grid part of RES energy produced.

TABLE1.
DETAILED DESCRIPTION OF THE PROPOSED INVESTMENTS

Location and name of buildings	**Province of Matera LS, LC, ITC2**
Number of buildings	3
Total surface (m2)	19.729
Current primary electricity energy consumption (MWh/year)	431
Current natural gas consumption(MWh/year)	1.211
Primary electricity energy savings (MWh/year)	872
Natural gas savings (MWh/year)	676
Energy savings %	100%
Natural gas savings %	56%
Average GHG emissions (tCO2e/m2/year)	321,27
Estimated CO2 reduction %	90%

IV. CONCLUSIONS

The proposed approach of TESA process based on LCA applied to a real case of school complexes, led to an interesting paradigm of energy districts in which, through a mix of technological solutions, Energy Efficiency and RES interventions can be rejoined into an unique investment. This allows to optimize the interaction between local energy generation and consumption.

REFERENCES

[1] G. Finnveden, MZ Hauschild, T Ekvall, J Guinee, R. Heijungs, S. Hellweg, A. Koehler, D. Pennington, S. Suh, "Recent Developments in Life Cycle Assessment", *J Environ Manage*, 91 (1): 1–21, 2009

[2] A. Björklund, "Life cycle assessment as an analytical tool in strategic environmental assessment. Lessons learned from a case study on municipal energy planning in Sweden", *Environ Impact Assess Rev* 32:82–87, 2012

[3] S. Lundie, GM Peters, PC Beavis, "Life cycle assessment for sustainable metropolitan water systems planning", *Environ Sci Technol* 38:3465–3473, 2004.

[4] E. Loiseau, G. Junqua, P. Roux, V. Bellon-Maurel, "Environmental assessment of a territory: an overview of existing tools and methods", *J Environ Manage* 112:213–225, 2012.

[5] M. Bagliania, A. Galli. V. Niccolucci, N. Marchettini, "Ecological footprint analysis applied to a sub-national area: The case of the Province of Siena (Italy)", *Journal of Environmental Management* 86,354–364, 2008.

[6] Nitschelm, L. et al. ,"Utility of spatially explicit LCA for agricultural territories", *Proceedings of the 9th International Conference on Life Cycle Assessment in the Agri-Food Sector*,2014.

ekoNET system architecture and service for environmental monitoring

Boris Pokrić, Srđan Krčo, Dejan Drajić,Maja Pokrić
DunavNET doo Novi Sad
Antona Čehova 1/2, 21000 Novi Sad, Serbia
e-mail: boris.pokric@dunavnet.eu, srdjan.krco@dunavnet.eu, dejan.drajic@dunavnet.eu, maja.pokric@dunavnet.eu

***Abstract*— The ekoNET system is developed for a real-time monitoring of air pollution and other atmospheric condition parameters such as temperature, air pressure and humidity. The ekoNET service is based on EB800 device integrating low-cost gas, Particluate Matters (PM) and meteorological sensors providing cost-efficient, simple to deploy, use and maintain solution targeted for the usage within the IoT domain of smart cities and smart enterprises. This paper gives an overview of the overall system architecture, ekoNET device, back-end cloud IoT infrastructure mapped to the IoT-A Architecture Reference Model (ARM), data handling and visualization engine as well as the application-level components and modules.**

Keywords: *Smart City, IoT, low-cost sensors, environmental monitoring, CoAP*

I. INTRODUCTION

The Internet of Things (IoT) concept envisages smart environments connecting with the citizens and shaping the cities around the world by offering smart services and concepts with aim to increase the quality of life in these cities. Connecting the environments with people requires an interface with rich user experience able to engage the users and present the relevant information that will fulfill the goal of the service.

Currently, more than 50% of people live in cities and the UN estimates that by year 2050 cities will be home to 70% of the world's population [1]. Consequently, significant effort is directed towards accommodating this growing trend utilizing smart city concepts and ideas and relying on ICT. Furthermore, the citizens expect more from the cities: to have better quality of life and to have detailed information about the city's environmental conditions. For example, although cities occupy only 3% of the world's geography, they generate about 80% of CO_2 emission. Additionally, being a "green" city is important for cities to attract tourists, investors and business and to provide additional information to citizens about the environmental conditions at different areas of a city. At the same time the enterprises are continuously trying to optimize the work processes and at the same time act in a socially responsible manner. To achieve these aspirations, remote monitoring of processes including hazardous gas levels at different locations of the facility, industrial safety, personal exposure and eco-friendly solutions play an important part, in particular in the oil and gas industry. In the developed countries there is a trend of lowering tolerance to the air polluting and all other factors that influence the environment. This goes in hand with improvement of monitoring services and putting in place polices, which as consequence for not following the regulations impose adequate penalties. Companies increasingly have a need to monitor environmental conditions and take corrective actions much before any inspections takes place in order to avoid strong sanctions.

Ultimately, the main goal is to raise an awareness of the community of importance of the environmental issues, in particular the air quality and related air pollution. This task is not simple and requires innovative methods and approaches in attracting large community that will participate and get engaged in the activities related to these issues. This work is part of the comprehensive study and in this paper we present some interesting results regarding ekoNET solution, while the complete results and integration with AR (Augmented Reality) platform will be published in [2].

The paper is organized as follows. In the section II principles of air quality monitoring in smart cities are presented. ekoNET service and system architecture are described in the section III an IV respectively. Section V concludes the paper.

II. AIR QUALITY MONITORING

Currently, the air pollution within the cities is monitored by networks of static measurement stations usually operated by the public authorities. These fixed stations are highly reliable and able to accurately measure a wide range of air pollutants. However, they are very large, expensive and require significant amount of maintenance. Subsequently, the extensive cost of acquiring and operating these stations severely limits the number of installations.

As a result of this problem, low-cost solid-state gas sensors have started to be used for measuring the pollutants in the atmosphere. One of the most popular types of these sensors, have an electrochemical reaction, when exposed to a specific gas. The gas concentration is determined by measuring either the sensor's output current or the resistance of the sensor's tin dioxide layer. These solid-state gas sensors are inexpensive, small, and suitable for mobile measurements.

In the recent years the activities related to environment monitoring are becoming more intensive and are in the

focus of many research projects. Special attention is devoted to projects targeting development of smart cities and smart societies in general. Distributed sensing systems for environmental parameters and portable and personal monitoring are emerging as a potential novel solution for data collection with enhanced with computational tools for a real time analysis. Projects such as the European SmartSantander project [3], the Japanese DOCOMO Project [4], and the Copenhagen wheel project [5] use portable sensors to measure a number of parameters including air pollution, UV, noise and/or meteorological conditions. The portable sensors are placed around the city and/or on vehicles or bicycles.

Also, the recent literature covers a number of devices for air pollution monitoring in urban and rural areas that utilize low cost sensors and wireless systems for transmission of measured data. A system that measures concentrations of gases, such as CO, NO_2, SO_2, and O_3 using semiconductor sensors is presented [6] where the measurement station is static and there is no possibility of remote data reading. In [7] the authors present a wireless distributed mobile air pollution monitoring system which is implemented and tested using the GPRS public network. The system utilizes city buses to collect pollutant gases such as CO, NO_2 and SO_2. In [8] the concept of a mobile monitoring system for chemical agents control in the air is presented (CO, CO_2, NO, NO_2 and VOC). The proposed system can be applied to measure industrial and car traffic air pollution. Data transmission uses the GPRS/EDGE radio link to transfer measurement results to the server with database. ekoNET system, presented in this paper, contains the greater number of sensors than other devices, and have ability to perform measurements both at a static location and as a mobile platform.

In paper [9] authors present an approach based on the mathematical statistics methods towards significant reduction of the number of necessary measuring points, as well as the number of required sensors while still providing reliable estimation of environment parameters across the monitored area. The paper shows that proposed solution works quite well in the areas with slowly-changing weather conditions. Solution is verified on the mobile platform mounted onto public transport vehicles and used for measurements of environmental parameters.

The environmental monitoring solution ekoNET aims to provide a simple and cost-effective solution that can be deployed both within the cities and enterprises wishing to monitor the air pollution and atmospheric parameters. This paper presents the overall system architecture encompassing the ekoNET device, back-end cloud infrastructure.

III. EKONET SERVICE

The ekoNET service is designed in such a way to provide a complete end-to-end solution for the environmental monitoring following the design concepts used within the IoT domain. The system comprises all necessary components namely: devices (EB800), back-end infrastructure and client applications (web and mobile).

The back-end infrastructure and the overall system follows the design principles of IoT architecture reference models such as IoT-A ARM so that the overall design methodology can be easily compared and potentially integrated to the other IoT platforms.

The exact functionality in terms of the environmental monitoring is defined with types of sensors used within the device itself. Currently, the system is designed to support two types of sensors: more accurate ones that can be used for air quality measurements and less accurate sensors that can be used to provide indications of the levels of the gases in the air. The devices with both variants can be used indoor and outdoor and it is planned to develop personal devices as well, which will be used to monitor personal exposure.

The device is designed in modular fashion, making it possible to use different sensor packs, adapted to suit different industries and associated use-cases, while the core processing and communication part of the device remains the same. This is of particular importance for the industries where specific requirements and regulations are in place.

The devices can be mounted on the public transportation vehicles (buses and trolleybuses) and on the trucks in coal mine in order to monitor air quality in different parts of coal mine. End users can query the system using a web or mobile application to get the real time measurements as well as locations of the vehicles. The EB800 devices are designed in a modular fashion enabling connection of different sensors according to the requirements. The central part of the device is the main board which block diagram is shown in Figure 1.

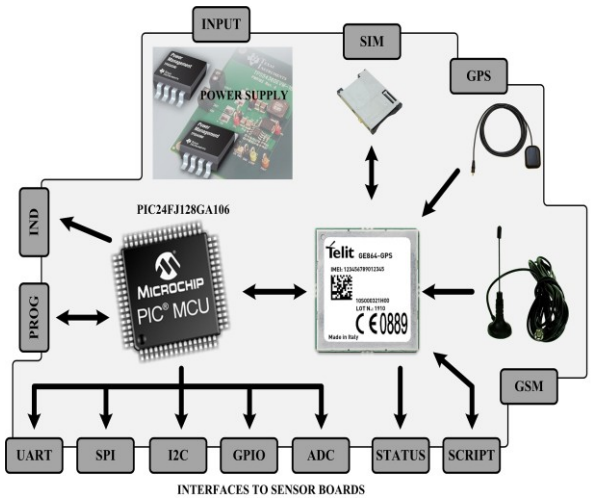


Figure 1. Block diagram of ekoNET device main board

It essentially provides the core functionality of the system such as GPRS/GPS connectivity as well as the digital and analog interfaces for the external sensor boards. All the components have been carefully selected in order to minimize the intrinsic noise generated by the

internal electronic circuitry and also to prevent any noise injection into the sensitive sensor driver circuitry. The main design philosophy that has been adopted it to connect all the sensor boards via the digital I2C or SPI interface. This decision has been undertaken in order to minimize the interference with the low-level analogue signals that sensors provide. In this way, each of the sensors has its own driver board with associated ADC that is directly connected to the main board via the I2C interface. In this way, the length of the analogue line (susceptible to the noise interference) is minimized and therefore the signal to noise ratio is greatly increased.

The sensors that are connected the main board are of different types, namely electrochemical gas sensors, atmospheric conditions sensors, particulate matter and noise.

The ekoNET device presented in this paper is equipped with the following atmospheric condition sensors:

- The sensor for atmospheric pressure measurements MPX4115 (15-115 kPa) [11]
- The sensor for temperature and humidity measurements CC2D23S (-40°C–123°C, 0-100%) [12]

Sensors used for the measurements of concentration of gases in the air are Alphasense's B4 family electrochemical types and infrared type IRC-AT for CO_2:

- CO_2-IRC-AT (0-5000ppm) [13]
- O_3-B4 (0-2ppm) [14]
- NO -B4 (0-20ppm) [15]
- NO_2-B4 (0-20ppm) [16]
- CO-B4 (0-50ppm) [17]

Alphasense B4 series sensors are intended for air quality monitoring in urban, rural and indoor areas, while D4 are industrial sensors. In this paper we are presenting results for urban air quality monitoring based on B4 sensors.

Furthermore, EB800 device is also equipped with the noise sensor measuring the sound pressure up to 105 dB.

As for the particulate matter monitoring Alphasense OPC-N1 [17]is used as particulate matter sensor, measuring PM1, PM2.5 and PM10, as well as measuring the particle size distribution in real time. This sensor solution uses SPI interface and can easily be attached to the main board. Alphasense OPC-N1 already has microcontroller for local signal processing and therefore it is connected to the main board via SPI connector.

Figure 2. shows the ekoNET device with mounted electrochemical gas sensors and packaged in the waterproofed box which makes this set-up suitable for the outdoor usage.

Figure 3. shows the EB800 components inside of the box with electrochemical gas sensors on the right, CO_2 at the top, main board in the middle of GPS/GPRS antenna at the left.

Figure 2. ekoNET device with low-cost electrochemical gas sensors mounted in the box

Figure 3. ekoNET device with low-cost electrochemical gas sensors

Each of the available devices is initially registered into the Resource Directory (RD) which stores the meta information about all resources and services within the IoT platform.

Subsequently, the collected data from the EB800 sensors are packaged into appropriate format and sent to the back-end cloud infrastructure via the mobile network (GPRS) channel utilizing Constrained Application Protocol (CoAP). The data is stored in the appropriate database which then service layer components use to provide the data to the applications.

The data is visualized in a real-time using the appropriate web or mobile application such as shown in Figure 4. The applications can also be used for the visualization of historical data which was gathered over longer period of time. In addition, as each ekoNET device comprises GPS module, the exact location of the sensor pack is known, so the network of environmental conditions can be built, pin pointing the "best" areas for the public, as well as indicating the areas where the pollutant levels are higher.

Figure 4. Web application for ekoNET data visualisation

IV. EKONET SYSTEM ARCHITECTURE

Figure 5. shows the high-level architecture of the ekoNET system mapped to the IoT-A Architecture Reference Model (ARM) [18]. The aim of the IoT-A ARM is to provide an architectural reference model for the interoperability of IoT systems, outlining principles and guidelines for the technical design of its protocols, interfaces and algorithms. Following the outlined principles ensures that the implemented and instantiated architecture complies with the standards related to the interoperability in terms of the protocols and functional specification of the building blocks of the resulting IoT system.

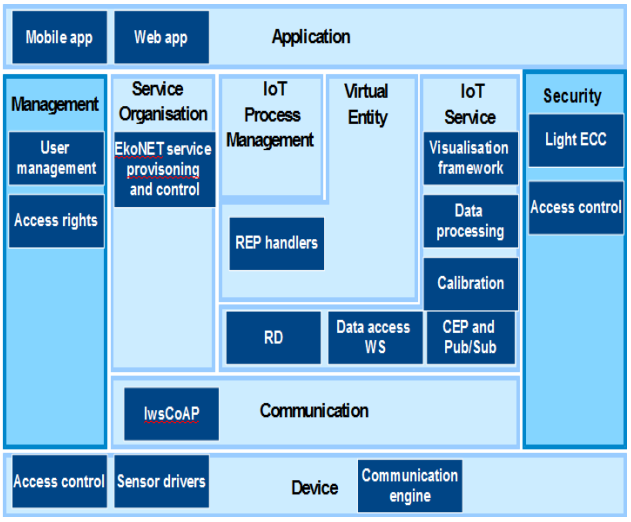


Figure 5. ekoNET system architecture mapped to IoT-A ARM

The ekoNET devices are connected to the ekoNET platform serving as the back-end cloud infrastructure. There are several variants of the ekoNET devices depending on the deployed location (indoor/outdoor), type of sensors used (air quality/safety) and usage (stationary/personal) but their deployment is same in the context of the connection to the cloud infrastructure. Each of the sensor has its Sensor driver board**.** Access to sensor data is controlled using the Access Control component which is also used for controlling the access of devices to the cloud infrastructure. This ensures that only authorized users can access the data as well as only authorized devices can access the cloud infrastructure. The collected data from the sensors is transmitted to the back-end infrastructure via the mobile network using the CoAP [19] utilizing the Communication engine component. CoAP is an application layer protocol designed to lower the complexity for the constrained networks but, also, to enable communication over the existing internet infrastructure. CoAP is a light-weight application protocol based on UDP that supports multicast requests, cashing and REST web services between the end-points, and is seen as a future protocol for IoT. CoAP is still work in progress of the IETF CoRE Working Group [20]. This system utilizes light weight secure CoAP (lwsCoAP) that is designed for secure data transfer using CoAP where the encryption is based on Elliptic Curve Cryptography (ECC) which significantly lowers the required computational effort for data encryption which is essential for the constrained IoT devices such as EB800. This component is also utilized on the server side to enable the creation of the secure communication channel between the devices and rest of the platform.

The cloud platform and associated building blocks enable the core features such as the permanent storage and access to the data (Data Storage and data access web services component) within the Data Server component). Search and discovery of resources and services is performed using the Resource Directory (RD) component. Complex Event Processing and Publish/Subscribe broker enable detection of complex events from multiple heterogeneous data sources. Once these events are detected, they are forwarded (published) to the users or components that consume them (subscribers). data cleaning and processing functionality (Data Server and Data Processing components) as well as a set of visualization widgets (Visualization Engine component). Calibration component is very important module in the context of the environmental monitoring sensors as it provides functionality for continuous correction of acquired data using the empirical models derived during the laboratory calibration phase. Data processing component is responsible for providing various levels of data analysis, used by the Visualization Framework for example (e.g. data interpolation, data averaging, medians, trends etc.). Visualization framework enables the application-level components to display the measured values utilizing set of widgets capable of showing real-time, historical and other statistical data sets. REP handlers provide the top-level entry points for accessing the EB800 devices provide that the access is allowed by the authentication and authorisation procedure. Service provisioning and control component is responsible for handling the creation of various services that can be provisioned on the system. For example a service can be created that provides temperature measurements every minute which only specific users are allowed to access. User management component provides functionality for

the creation of users, access rights based on the access rights rules defined for the entire platform.

On top of the architecture stack there are web and mobile applications that are used to access then services provided by the platform. Furthermore, additional web applications are used for the administrative purposes.

V. CONCLUSION

In this paper we have presented the environmental monitoring service ekoNET, based on the low-cost electrochemical gas sensors. The ekoNET system is developed for a real-time monitoring of air pollution and other atmospheric condition parameters and is intended for the usage within the IoT domain of smart cities and smart enterprises. Overview of the overall system architecture, ekoNET device, back-end cloud IoT infrastructure mapped to the IoT-A Architecture Reference Model (ARM), data handling and visualization engine as well as the application-level components and modules are given and discussed in details. In the future work ekoNET will be integrated with AR Genie, augmented reality platform. By extending the AR Genie platform with the ekoNET IoT service, we will able to demonstrate usage of real-time environmental data within AR mobile applications.

ACKNOWLEDGMENT

This paper describes work undertaken in the context of the SocIoTal project (http://sociotal.eu/). The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement n° CNECT-ICT- 609112.

REFERENCES

[1] http://www.who.int/gho/urban_health/situation_trends/urban_population_growth_text/en/

[2] Boris Pokrić, Srđan Krčo, Dejan Drajić, Maja Pokrić, Vladimir Rajs, Zivorad Mihajlović, Petar Knežević, Dejan Jovanović, "Augmented Reality enabled IoT services for environmental monitoring utilising serious gaming concept" accepted for publication in *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, Vol. 6, No. 1 March 2015

[3] http://www.smartsantander.eu/

[4] http://www.nttdocomo.com/pr/2009/001461.html

[5] http://senseable.mit.edu/copenhagenwheel/

[6] N. Kularatna, B. Sudantha, "An Environmental Air Pollution Monitoring System Based on the IEEE 1451 Standard for Low Cost Requirements", *IEEE Sensors Journal*, vol. 8, no. 4, pp. 415-422, Apr. 2008.

[7] A. R. Al-Ali, Member, IEEE, Imran Zualkernan, and Fadi Aloul, Senior Member, IEEE, "A Mobile GPRS-Sensors Array for Air Pollution Monitoring" *IEEE Sensors Journal*, vol. 10, no. 10, pp. 1666-1670, Oct. 2010

[8] Ryszard J. Katulski, Jacek Namieśnik, Jacek Stefański, Jarosław Sadowski,Waldemar Wardencki, Krystyna Szymańska, "Mobile monitoring system for gaseous air pollution". *Metrology and Measurement Systems*, vol. 16, no. 4, pp. 667-682, Dec 2009

[9] Vladimir Rajs, Vladimir Milosavljević B, Živorad Mihajlović, Miloš Živanov, Srđan Krčo, Dejan Drajić, Boris Pokrić, "Realization of Instrument for Environmental Parameters Measuring", *Electronica ir Elektrotechnika*, Vol. 20. No. 6 pp-61-66, June 2014

[10] http://www.freescale.com/files/sensors/doc/data_sheet/MPX4115.pdf

[11] http://www.sensirion.com/fileadmin/user_upload/customers/sensirion/Dokumente/Humidity/Sensirion_Humidity_SHT7x_Datasheet_V5.pdf

[12] Alpha Sense Gas Sensor Datasheets. [Online] http://www.alphasense.com/WEB1213/wp-content/uploads/2014/01/IRC-A1.pdf

[13] Alpha Sense Gas Sensor Datasheets. [Online] http://www.alphasense.com/WEB1213/wp-content/uploads/2013/11/O3B4.pdf

[14] Alpha Sense Gas Sensor Datasheets. [Online] http://www.alphasense.com/WEB1213/wp-content/uploads/2013/11/NOB4.pdf

[15] Alpha Sense Gas Sensor Datasheets. [Online] http://www.alphasense.com/WEB1213/wp-content/uploads/2013/11/NO2B4.pdf

[16] Alpha Sense Gas Sensor Datasheets. [Online] http://www.alphasense.com/WEB1213/wp-content/uploads/2013/11/COB4.pdf

[17] http://www.apollounion.com/Upload/DownFiles/Upload_DownFiles_OPC-N1.pdf

[18] IoT-A Architecture Reference Model, http://www.iot-a.eu/public/public-documents/copy_of_d1.2/view

[19] CoAP Specification, https://datatracker.ietf.org/doc/rfc7252/

[20] http://datatracker.ietf.org/wg/core/charter/ IETF CoRE Working Group [Last accessed April 2013]

Software Module for Integrated Energy Dispatch Optimization

Marko Batić, Nikola Tomašević, Sanja Vraneš
School of Electrical Engineering, University of Belgrade, Institute Mihajlo Pupin, Belgrade, Serbia
marko.batic@pupin.rs, nikola.tomasevic@pupin.rs, sanja.vranes@pupin.rs

***Abstract* — Continuous increase of global energy demand combined with the lack of conventional energy sources resulted in rise of operational costs for energy infrastructures. This calls for systematic and comprehensive Energy Management (EM) approaches that are able to take the advantage of increasingly used Renewable Energy Sources (RES) as well as to provide for more efficient operation of existing multi-carrier energy infrastructures. The presented work focuses on the development of an extensive, flexible and modular simulation tool for the analysis and optimisation of such energy systems based on extended Energy Hub concept. The tool was developed as a software module which allows for both energy planning and operation scenarios within multi-carrier environment. The prototype of the analysis and optimisation tool was first developed in MATLAB® and subsequently a corresponding software module was implemented in Java® environment. Technical details and challenges associated with the implementation of software module were elaborated in detail after which the applicability of developed software is highlighted and finally an example of use-case is given.**

I. INTRODUCTION

The trends of global energy conservation seek for energy savings through various means such as inclusion of Renewable Energy Sources (RES) as well as better utilization of conventional energy sources through improved efficiency of ongoing energy intensive processes. Apart from environmental aspects, these energy saving activities are also coupled with high economical impact since any savings in energy can be easily associated with tangible monetary repercussions. In order to reconcile these two there is a need for a comprehensive Energy Management (EM) solution that will be able to economically evaluate integration of RES in long term as well as to provide efficient operation of existing infrastructures on daily basis. The first objective requires careful consideration of various hybrid energy infrastructures in order to successfully tackle the stochastic nature of RES and provide continuous power supply to the end user. On the other hand the second objective entails satisfaction of the user demand in the most efficient and hence the most economic way.

This paper therefore especially focuses on the development of an EM software tool used partly for long term economical appraisal related to small-scale hybrid renewable energy systems, suitable for supplying the residential users with clean but also reliable source of energy and partly as an energy dispatch module that will be able to integrate on existing Building Management Systems (BMS) and conduct smart energy management strategies.

A. State of the Art

An exhaustive research of the state of the art in the area of RET simulators and optimizers with the similar objective was performed and the most prominent were recognized. Starting with the NREL [1], a computer model called HOMER [2] was developed, which deals with evaluation of design options for large-scale off-grid and grid-connected power systems for remote, stand-alone, and distributed generation applications. HOMER is presently available as fully commercial tool. Another tool called REopt [3] serving as an energy planning platform that offers concurrent, multiple technology integration and optimization capabilities to help clients evaluate potential savings and energy performance objectives was developed in the same laboratory. Following is the Hybrid2 [4] software package, developed in RERL [5], representing a tool performing long term performance and economic analysis on a wide variety of hybrid power systems. Furthermore, software for clean energy project analysis, called RETScreen [6] was developed with objective to provide for a decision support tool for RET deployment. Finally, a RET simulation and optimization software called iHOGA [7], based on utilization of genetic algorithms, was also analyzed.

However, none of the aforementioned applications provides a light-weight, web-enabled and easy to integrate tool serving both for energy planning decision support as well as optimal operation of existing energy infrastructures.

B. Selected approach

The objective of this paper is to elaborate on the software module built on the extended Energy Hub (EH) concept described in [8] and [9]. This concept takes the advantage of existing Supervisory Control and Data Acquisition (SCADA) systems and additional metering, if necessary, to provide energy dispatch optimization of complex multi-carrier energy infrastructures. The original EH concept offers modelling of energy flows from different energy carriers while satisfying the requested user demand [10]. The concept leverages on the conversion potential of a specific, constrained, domain referred as Hub which serves as a point of coupling between existing energy supply infrastructures and energy end use. The Hub basically represents a set of energy converters and/or storages which is responsible for delivering required energy by taking into consideration different conversion and/or storage options while meeting a desired optimization criterion. So far, many aspects of the EH have been thoroughly elaborated, thus emphasizing optimization potential of the concept owing

to its flexible modelling framework, diverse technologies and wide range of energy carriers [11][12]. The latest research efforts even considered generalization of this concept by introducing renewable energy sources, which was first mentioned in [13]. However, considering that EH concept basically performs optimization of supply side, without affecting the desired energy demand, the concept has been further improved by including additional, complementary, optimization of the demand side, which proved to create space for further energy cost savings. This implied utilization of the well-known concept of demand side management (DSM), which consists of various techniques for modifying the energy end use profile, i.e. the demand side. However, it should be emphasized that any further savings, compared to EH approach, require certain level of compromise from the user (changing the time schedules of equipment, reducing the demand etc.). Nevertheless, this is perfectly aligned with current trends in energy supply as more and more energy providers offer significant economic benefits if, in return, the end user complies with some energy end use constraints (reducing loads in peak hours, improving power factor etc.).

The remainder of the paper starts with the Section II describing the existing EH concept and its modelling framework. The Section III briefly elaborates on the modelling and development aspects of the optimization engine prototype. The actual software module is elaborated in Section IV, where all technical details are revealed. Finally, the paper is concluded and the impact of the developed software is discussed in Section V.

II. Energy Dispatch Optimization Concept

As introduced in previous sub-section, the selected approach for development of energy dispatch optimization framework is based on integrated optimization of both supply and demand side of so called energy Hub. Hence, a mathematical representation and theoretical basis for optimization of such a Hub are given in the following.

From the modelling perspective, Hub is represented as a matrix which includes, in the most generic case, elements which enable conversion of various supply energy carriers to satisfy different load types. Furthermore, since physical Hub may also take into account different types of storages, each energy carrier is associated with its storage unit which acts as energy buffer at the cost of storage efficiency, and the corresponding storage matrix is considered as well. This is depicted in the Hub's conceptual schemata shown in Figure 1. As depicted, the power input comprising of conventional (P) energy sources, such as electricity power grid, natural gas, district heating, fossil fuels etc., is supplied to the Hub. Apart from the power input (P), a vector comprising of all local energy production (R), such as photovoltaic, wind turbines for electricity and/or solar thermal and geothermal for thermal energy, is also considered as Hub's input. The input power is then transformed using the conversion elements (C), allowing for conversion from electrical towards thermal energy and vice versa, and/or energy storages (Ė), such as batteries, ultra capacitors, fuel cells for electricity or boilers and phase changing materials for thermal energy, while taking into account the storage efficiencies depicted with coupling matrix (S). Passing through the Hub, depicted by the conversion and/or storage matrix, power from the supply side is fed to

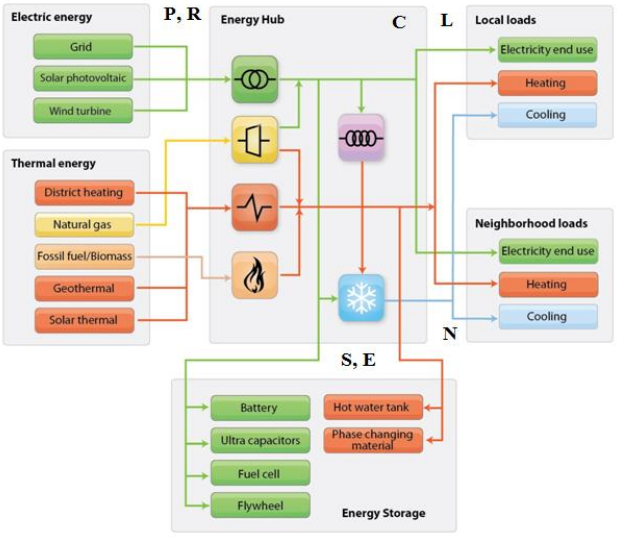


Figure 1. Energy Hub concept

the demand, loads (L), typically represented with electricity and heating/cooling loads. The output of the Hub is also complemented with the vector depicting neighbourhood loads (N), while preserving the same distribution between electricity and heating/cooling loads, which allow the Hub to feed the surplus (export) of energy towards the neighbourhood, which is considered to be another similar entity or a piece of power infrastructure. Finally, the complete Energy Hub model equation, defined in [11], is given in the following:

$$(L+N) = C(P+R) - S\dot{E} = [C \quad -S]\begin{bmatrix} P+R \\ \dot{E} \end{bmatrix}$$

Considering the flexibility and generality of such modelling approach, a Hub concept can be applied to an entity ranging from single residence up to an entire city or country. Once all the matrices and vector parameters are defined, according to described model, an optimization problem is hence formulated and a corresponding solver was applied. Based on the research efforts related to the development of integrated energy dispatch optimization concept, described in more detail in [14], the following are key features of the actual software module prototype as well as mature application.

III. Optimization Engine Prototype Development

A prototype of the software module for the analysis and optimisation of multi-carrier energy systems described as Hubs has been implemented in MathWorks®/MATLAB® environment based on the defined mathematical model. Given the formulation of the Hub, the optimization problem depicting energy dispatch, may be represented as Linear Program (LP). Although there is an abundance of solvers dealing with that kind of optimisation problem, it was decided that IBM® ILOG® CPLEX® Optimizer should be used as it is one of the industry standards. It requires the following input data:

- *f*: vector for the definition of the objective function
- *Aineq*: matrix for inequality constraints
- *bineq*: vector for inequality constraints
- *Aeq*: matrix for equality constraints

- *beq*: vector for equality constraints

The major functionality of the developed MATLAB code is to define the system data in matrix and vector form and translate this information into the input matrices and vectors required by the employed solver. This process can be efficiently performed by exploiting the powerful matrix and vector manipulation routines of the MATLAB engine. The implemented MATLAB code has the following major functional sections.

1) **Data Input**: The energy hub model of the physical system is defined by defining the model matrices, i.e. the input storage matrix *Sqin*, input dispatch matrix *Fin*, conversion matrix *C*, output dispatch matrix *Fout*, output storage matrix *Sqout* as described in Section 2. At a second stage other parameters needed for the formulation of the optimisation problem (such as load data, price data, etc.) are defined as time series or read in from corresponding files.

2) **Solution:** After the matrix and vector parameters defining the optimisation problem have been formulated the solver routine is called to execute the optimisation.

3) **Result Processing:** If an optimal solution has been found by the optimisation function, the optimisation results are processed for display purposes.

IV. Optimization Engine as Software Application

A. Technology considerations

After the development of integrated energy dispatch optimization engine prototype in MATLAB environment, together with CPLEX Optimizer, the objective was proceed with its development within such environment that would enable integration of the developed software module with existing BMS. Having this in mind and considering that majority of the BMS components are mainly developed under the object oriented paradigm, the choice was made to opt for Java™ framework. More precisely, the development of business logic components was done using Java programming language whereas the corresponding Web technologies were used to enable remote access to the engine. However, considering the developed MATLAB prototype, there were two alternatives for integration within Java environment:

- To generate Java classes out of the MATLAB source code using the MATLAB Builder™ JA - Then Java classes can be integrated into Java applications and deployed royalty-free to desktop computers or web servers that do not have MATLAB installed using the MATLAB Compiler Runtime (MCR) that is provided with MATLAB Compiler™.
- Or, to develop native Java application from scratch offering the same functionalities as developed prototype.

Naturally, both approaches have their advantages and disadvantages, as explained in the following.

The first one takes the advantage of the implemented source code and does not require in-depth knowledge of the underlying algorithms. However, it requires additional pre- and post- processing of the information stored in the generated Java classes. Also, considering the fact that MATLAB prototype is using external optimizer (CPLEX), which is dependent on the programming environment, it would be necessary to break the MATLAB source into two parts, i.e. the one before the call of CPLEX functions and the other one after, to provide for corresponding CPLEX library for Java. However, using CPLEX in JAVA environment is quite different than in MATLAB, where it is required only to replace the call towards native MATLAB functions with call towards CPLEX functions using the same arguments. This means that even if the MATLAB source was used, although wrapped in object oriented paradigm, access towards CPLEX would have to be completely changed in Java environment. Using CPLEX in Java environment is, however, a bit more complex and requires utilization of a specialized Application Program Interface (API). This allows Java application to call CPLEX directly, through the JNI (Java Native Interface). The Java interface is built on top of ILOG Concert Technology for Java and supplies a rich functionality allowing you to use Java objects to build an optimization model. This concept is depicted in Figure 2. However, this would require additional coding to adapt both inputs and outputs coming from the MATALB code, in order to be able to run CPLEX optimization engine within Java environment.

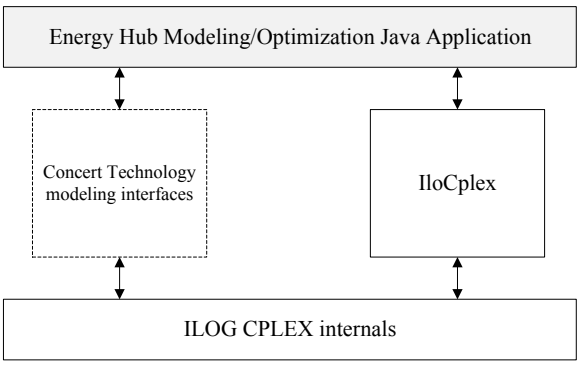


Figure 2. ILOG Concert Technology for Java (taken from http://www.cs.cornell.edu/)

On the other hand, the second approach of pure Java application requires in-depth knowledge of underlying algorithms in order to perform semantic translation, unlike in the previous case where it was possible to perform translation solely on the syntax level. However, since the use of CPLEX API for Java is inevitable in any case, developing application in pure Java is more favoured approach for many reasons. It represents a more streamlined approach with many advantages when it comes to the model scalability, easy maintenance of the code and, the most importantly, shorter execution times. Therefore, the rest of this section aims at providing an insight to the implementation of optimization engine as pure Java application.

B. Java application development

The optimization engine is basically a Java application based on Java Enterprise Edition 7 (Java EE 7). For the development and implementation of this application an Integrated Development Environment (Eclipse Java EE IDE - Juno Service Release 2) was utilized. Furthermore, having in mind that one of the main objectives was also to enable remote access to the optimization engine results, further developments were conducted in the direction of developing an online tool easily accessible at any time or place from a web browser. This suggested use of a thin

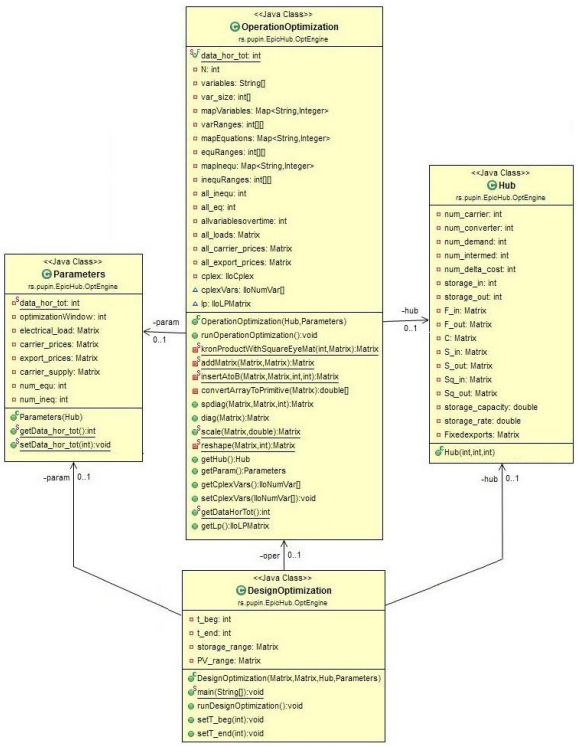


Figure 4. Comparison of total dispatch costs distribution

client and three-tier software system architecture, which has proven to be an immensely useful for the enterprise application development. Therefore, an Application Server (Apache/Tomcat 7.0) was employed as well.

As regards external components (apart from Java System Library) that are necessary for engine implementation, they are as following:

1) CPLEX API – represents a library which contains CPLEX Concert Technology for Java, thus offering API that includes modelling facilities to allow the programmer to embed CPLEX optimizer in Java application. Therefore, it provides a set of interfaces and classes that enable typical CPLEX features such as creating a model, solving the model, querying results after solving, and handling error conditions.

2) CPLEX core engine– represents implementation of CPLEX optimizer in a dynamic linked library which is needed to be able to run Java applications that use CPLEX. In other words, this represents a light weight approach where there is no need for the installation of the overall ILOG CPLEX Optimization Studio for running CPLEX applications.

3) Matrix manipulation – represents an open source, pure Java library that provides Linear Algebra primitives (matrices and vectors) and algorithms. Given the fact that the overall mathematical modelling and representation of Energy Hub is done using matrices use of Linear Algebra library was imperative.

The first two components were employed based on the previous decision to use CPLEX for solving our LP and MILP problems. Although there are other, open source solver solutions, the CPLEX was picked owing to its unprecedented performance and availability through Academic Licence. The last component, La4J, was selected based on a search for Java library for Linear

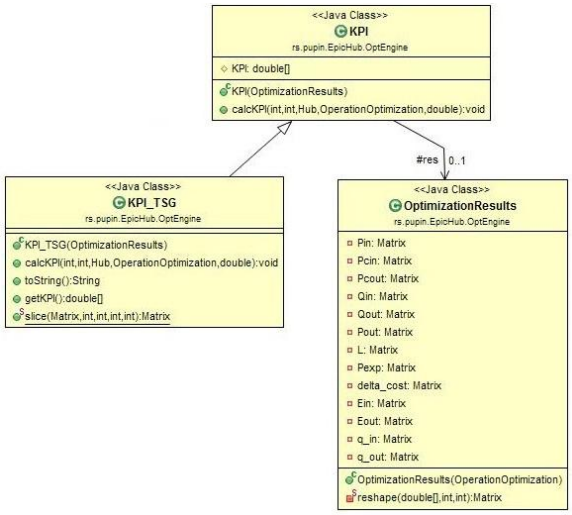


Figure 3. Comparison of total dispatch costs distribution

Algebra that provides manipulation with extremely large matrices. This implies utilization of concept of sparse matrices which enables viable manipulation of matrices with over 100k rows and columns. La4J was picked out of the group of several libraries that were tested such as: Commons_math, Colt, Ojalgo, JBlas, EjML, MtJ, Jeigen/Eigen and La4J.

When it comes to the implementation of the optimization engine as Java application, it was done using 7 classes as depicted in the corresponding UML diagrams in Figure 4 and Figure 3. The first one depicts classes that were used for actual optimization whereas the second one represents classes responsible for post-processing of the optimization results and calculation of corresponding Key Performance Indicators (KPIs) used for evaluation of different alternatives. Hence, Figure 4 depicts classes that implement the following key features:

- Hub operation optimization – implements the energy dispatch optimization of a particular Hub configuration. It uses the mathematical representation of Hub and corresponding parameters to model its behaviour over a given time span as LP problem. Furthermore, in order to be able to run the CPLEX optimizer, modelling was performed using the previously mentioned API and the corresponding interfaces as in following:
  - IloNumVar – modelling variables
  - IloRange – ranged constraints
  - IloObjective – optimization objective
  - IloNumExpr – expression using variables.
- Hub design optimization – features design optimization as multiple concentric loops going over different combinations of corresponding energy asset alternatives (namely different energy sources and storages) which are then used to run corresponding energy dispatch optimization.
- Hub modelling – models the topology of Energy Hub by defining corresponding matrices and necessary parameters. One of features of developed software module is ability to generate artificial data for various time variable input parameters (production from renewables, energy pricing and load profiles) which enables testing of

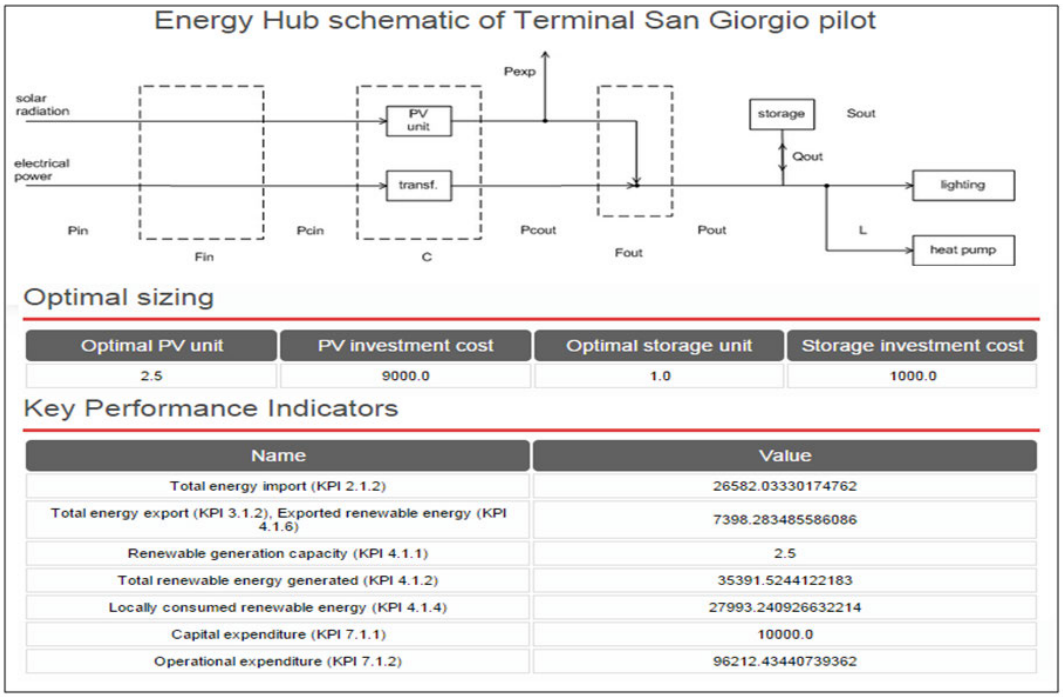


Figure 5. Example result from the optimization software module

the module itself but also provides opportunity to conduct sensitivity analysis of the proposed design solution.

V. CONCLUSION

This paper presents the work performed to conceptually develop and implement a software module for the analysis, simulation and energy dispatch optimisation of multi-carrier hub systems. The employed mathematical formalisation of the system equations allows for a modular approach which can be exploited at the software development stage to provide a tool that is efficient and flexible in its capability to model systems of different sizes and complexity. Furthermore, the presented tool enables the full numerical computation of analytical results starting from the physical data and parameters describing the pilot and will form a key part of the functionalities featured in the overall energy management platform. Starting from the prototype developed in MATLAB environment, the tool has finally grown to a high maturity level as Java application, suitable for integration with existing energy management software. Figure 5 contains a piece of user interface of web instance of the software module which is used to asses different sizing options for a given pilot. The figure is depicting a use case of the Terminal San Giorgio, part of the Genoa harbour, along with the recommendations for the optimal retrofit of energy infrastructure. Given the meteorological conditions of the terminal as well as its daily demand, a choice was made to opt for the installation of both PV panels and corresponding storage units. Also, a list of the key performance indicators for the suggested configuration is given enabling easier benchmarking with non preferred options. Further work will be focused on the development of REST-full services that will offer these results over machine readable interface and allow for easier integration with EM systems.

ACKNOWLEDGMENT

The research presented in this paper is partly financed by the European Union (FP7 EPIC-HUB project, Pr. No: 600067), and partly by the Ministry of Science and Technological Development of Republic of Serbia (SOFIA project, Pr. No: TR-32010).

REFERENCES

[1] National Renewable Energy Laboratory (NREL) – U.S. Department of Energy, http://www.nrel.gov/

[2] HOMER - Energy Modeling Software for Hybrid Renewable Energy Systems, http://homerenergy.com/index.html

[3] Renewable Energy Optimization Tool (REopt), NREL, Available:http://www.nrel.gov/tech_deployment/tools_reopt.html

[4] Hybrid2 - The Hybrid Power System Simulation Model, http://www.ceere.org/rerl/projects/software/hybrid2/

[5] Renewable Energy Recourses Laboratory (RERL),http://ywang.eng.uci.edu/About%20RERL.htm

[6] RETScreen, Natural Resources Canada, http://www.retscreen.net/ang/home.php

[7] iHOGA - improved Hybrid Optimization by Genetic Algorithms, http://www.unizar.es/rdufo/hoga-eng.htm

[8] Marko Batic, Nikola Tomasevic, Sanja Vranes, "*Integrated Energy Dispatch Approach Based on Energy Hub and DSM*" ICIST 2014, 4rd International Conference on Information Society and Technology, ISBN: 978-86-85525-14-8, pp. 67-72, Kopaonik, 09-13.03.2014.

[9] Nikola Tomasevic, Marko Batic, Sanja Vranes, "*Genetic Algorithm Based Energy Demand-Side Management*" ICIST 2014, 4rd International Conference on Information Society and Technology, ISBN: 978-86-85525-14-8, pp. 61-66, Kopaonik, 09-13.03.2014.

[10] P. Favre-Perrod, M. Geidl, B. Klöckl and G. Koeppel, "*A Vision of Future Energy Networks*", presented at the IEEE PES Inaugural Conference and Exposition in Africa, Durban, South Africa, 2005.

[11] M. Geidl, **"***Integrated Modeling and Optimization of Multi-Carrier Energy Systems*", Ph.D. dissertation, ETH Diss. 17141, 2007.

[12] B. Klöckl, P. Stricker, and G. Koeppel, "*On the properties of stochastic power sources in combination with local energy*

storage", CIGRÉ Symposium on Power Systems with Dispersed Generation, Athens, Greece, 13-16 April, 2005.

[13] M. Schulze, L. Friedrich, and M. Gautschi, "*Modeling and Optimization of Renewables: Applying the Energy Hub Approach*", IEEE conference, July 2008.

[14] EPIC-Hub Project Deliverable, "*D2.2 Energy Hub Models of the System and Tools for Analysis and Optimization*", Lead contributor Eidgenoessische Technische Hochschule Zuerich (ETH), 2014.

Experimental Evaluation of Growing and Pruning Hyper Basis Function Neural Networks Trained with Extended Information Filter

Najdan Vuković*, Marko Mitić**,

Milica Petrović**, Jelena Petronijević**, Zoran Miljković**

* University of Belgrade-Faculty of Mechanical Engineering/Innovation Center, Belgrade, Republic of Serbia

** University of Belgrade-Faculty of Mechanical Engineering/Production Engineering Department, Belgrade, Republic of Serbia

{nvukovic, mmitic, mmpetrovic, jpetronijevic, zmiljkovic}@mas.bg.ac.rs

***Abstract*— In this paper we test Extended Information Filter (EIF) for sequential training of Hyper Basis Function Neural Networks with growing and pruning ability (HBF-GP). The HBF neuron allows different scaling of input dimensions to provide better generalization property when dealing with complex nonlinear problems in engineering practice. The main intuition behind HBF is in generalization of Gaussian type of neuron that applies Mahalanobis-like distance as a distance metrics between input training sample and prototype vector. We exploit concept of neuron's significance and allow growing and pruning of HBF neurons during sequential learning process. From engineer's perspective, EIF is attractive for training of neural networks because it allows a designer to have scarce initial knowledge of the system/problem. Extensive experimental study shows that HBF neural network trained with EIF achieves same prediction error and compactness of network topology when compared to EKF, but without the need to know initial state uncertainty, which is its main advantage over EKF.**

I. INTRODUCTION

Radial basis function (RBF) neural networks are among the most used single layered neural networks [1, 2]. They are popular choice made by many engineers for modeling of complex real world problems [1-7].

In this paper, we develop and evaluate sequential learning algorithm based on Extended Information Filter (EIF) of special class of generalized RBF neural networks with Gaussian basis function that allows: (i) growing of neurons, (ii) pruning of neurons, and (iii) scaling of local input dimensions. Compared to the RBF, the hyper basis function (HBF) has different scale for each dimension of the input vector. Research results [1, 2] have shown that HBF generates neural network with same accuracy as RBF or even higher accuracy than RBF, but with less number of basis functions [2, 8-12]. Learning algorithm for HBF network learning algorithm is sequential and during sequential learning using Extended Information Filter (EIF) it changes number of HBF neurons according to predefined optimality criteria. In this paper, growing and pruning ability of HBF neural network is founded on the original contribution of the neuron's significance, introduced by Huang et al. in series of papers [13, 14], modified by Bortman and Aladjem [15] and generalized in [2], with introduction of Gaussian mixture model (GMM) for modeling of complex input densities [16].

This paper is structured as follows: in the second part of the paper we provide basic information related to intuition behind HBF network; in the third part the main learning algorithm is presented. Experimental results are presented in the fourth part, while concluding remarks are given in the final part.

II. HYPER BASIS FUNCTION NEURAL NETWORKS

The general mathematical form of RBF neural network is given as:

$$\mathbf{y}_i = \mathbf{f}\left(\mathbf{x}_i\right) = \sum_{j=1}^{J} w_j g_j\left(\mathbf{x}_i, \boldsymbol{\mu}_j, \sigma_j\right) \tag{1}$$

where $g_j(\cdot,\cdot,\cdot)$ stands for the j-th basis function; $\mathbf{x}_i \in \mathbb{R}^{n_x}$ is the input vector in the RBF neural network and n_x stands for the number of the input dimensions, $\boldsymbol{\mu}_j \in \mathbb{R}^{n_x}$ is the center of j-th basis function (also referred as the prototype vector); w_j is connecting weight of the j-th basis function, and σ_j is the spread ($\sigma_j \in \mathbb{R}^1$). Hyper basis function uses the Mahalanobis-like distance and it is given in the following form [1, 2]:

$$\begin{aligned} g_j\left(\mathbf{x}_i, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right) &= \exp\left(-0.5\left\|\mathbf{x}_i - \boldsymbol{\mu}_j\right\|_{\boldsymbol{\Sigma}_j}^2\right) = \\ &\exp\left\{-0.5\left(\mathbf{x}_i - \boldsymbol{\mu}_j\right)^T \boldsymbol{\Sigma}_j \left(\mathbf{x}_i - \boldsymbol{\mu}_j\right)\right\} \end{aligned} \tag{2}$$

where $\|\cdot\|_{\boldsymbol{\Sigma}_j}^2$ is Mahalanobis-like norm weighted with positive definite square matrix $\boldsymbol{\Sigma}_j$. Weighting matrix $\boldsymbol{\Sigma}_j$ makes similarity between the input vector $\mathbf{x}_i$ and j-th center vector $\boldsymbol{\mu}_j$ invariant to scaling and local orientation of the data [8-12]. The Mahalanobis-like distance can be seen as a generalization of the Euclidian distance for

similarity measure. The general case ($\mathbf{\Sigma}_j$ is full matrix) provides flexibility and bigger number of parameters to optimize but it results in severe over-fitting to data [2]. Therefore, in HBF network each HBF neuron has a unique diagonal weighting matrix $\mathbf{\Sigma}_j$, with varying size and restricted orientation, i.e. $\mathbf{\Sigma}_j = diag\left(1/\sigma_1^2, 1/\sigma_2^2, ..., 1/\sigma_{n_x}^2\right)$ [2].

This parameterization provides a trade-off between two extremes: on one hand, we have case where local scaling of data is not allowed, and on the other hand, the case with high degree of freedom of the RBF neural network model. Therefore, with diagonal weighting matrix we are trying to capture additional information. Let us show how important is to have the ability of local scaling of the data, especially for modeling of nonlinear dynamical systems. Let us observe NARX model [2, 8]

$$\begin{aligned}\mathbf{y}(t) = f\big(&\mathbf{y}(t-1),...,\mathbf{y}(t-l_y),...\\ &\mathbf{u}(t-1),...,\mathbf{u}(t-l_u)\big)+\mathbf{e}(t)\end{aligned} \tag{3}$$

where $\mathbf{u}(t)$, $\mathbf{y}(t)$ and $\mathbf{e}(t)$ represent system input, output and noise variables (respectively), l_y and l_u are maximum lags of the input and output, and $f(\cdot)$ is some unknown nonlinear function. Let us form the input vector into neural network given by:

$$\mathbf{z} = \left[\mathbf{z}_1(t),...,\mathbf{z}_l(t)\right]^T, \tag{4}$$

where $l = l_y + l_u$ and elements of $\mathbf{z}$ are:

$$\mathbf{z}_k = \begin{cases} \mathbf{y}(t-k),\ k = 1,...,l_y \\ \mathbf{u}\left(t-\left(k-l_y\right)\right),\ k = l_y+1,...,l_y+l_u \end{cases} \tag{5}$$

Now, without loss of generality, let us assume that input $\mathbf{u}(t)$ and output $\mathbf{y}(t)$ are bounded in $\left[\underline{u},\overline{u}\right]$, $\left[\underline{y},\overline{y}\right]$ and let us define associated ranges as $r_u = \left[\underline{u},\overline{u}\right]$ and $r_y = \left[\underline{y},\overline{y}\right]$. Two common cases can be distinguished: (i) $r_y \ll r_u$ (r_u is much greater than r_y) and (ii) $r_y \gg r_u$ (r_u much less than r_y). In the first case, influence of lagged input variables $\mathbf{u}(t)$ may be exaggerated while the role of states $\mathbf{y}(t)$ may be downplayed; in the second case the opposite is valid: the role of system variables $\mathbf{y}(t)$ will be exaggerated and the role of system inputs $\mathbf{u}(t)$ will be downplayed. This problem is commonly encountered in black box modeling of nonlinear dynamical systems when designers has no prior knowledge of the influence input dimensions have on the output.

It goes without saying that RBF neuron is especially vulnerable to this problem. Some parts of this problem may be solved by normalizing the data; however, the major part will still be unsolved. Having this in mind, the HBF neuron provides natural extension of the RBF neuron that provides us with ability to locally scale input data.

III. EXTENDED INFORMATION FILTER FOR GROWING AND PRUNING HBF NETWORK TRAINING – EIF-HBF-GP

A. Extended Information Filter

EKF is a widely established as one of the most successful learning algorithms for neural networks [1-3, 13-15, 17]. Although theory behind Kalman filtering and filtering in information space is well known [3], to best of our knowledge, EIF has not been applied for machine learning of neural networks, which is the reason why we decided to test its performance. The first attempt in this direction is undertaken in references [3-5] for RBF neural networks without ability to change network topology while learning. In this paper, we test performance of EIF for HBF-GP network training.

The attractiveness of EIF-based sequential learning algorithm is in parameterization of Gaussian distribution. Namely, instead of mean and covariance, information filter parametrizes Gaussian distribution with information vector and information matrix, which is defined as inverse of covariance, i.e.

$$\mathbf{I} = \mathbf{P}^{-1} \tag{6}$$

Now, the covariance matrix tells us how much we do not know about our system; bigger **P** means more uncertainty. If all elements of **P** are large, than it means that, we have no knowledge of our system's initial state. Similarly, all elements of **P** are small, than it means that we have all available information about initial state of our system (needless to say that we as engineers are aware that this ideal situation is not possible). Now, from computational perspective the problem arises when we have no initial knowledge of system's state; the problem is how to "tell" the computer this information. This is why we perform estimation in information space; namely, when we invert covariance matrix it is possible to tell the computer that we have little or no knowledge of system's initial state, which means that out lack of knowledge may be represented in the following symbolical mathematical form:

$$\begin{aligned} \mathbf{P} = \mathbf{0} \quad &\Rightarrow \quad \mathbf{I} = \infty \\ \mathbf{P} = \infty \quad &\Rightarrow \quad \mathbf{I} = \mathbf{0} \end{aligned} \tag{7}$$

In the first case, all elements of covariance matrix **P**(i,j) are small numbers, and all elements of **I**(i,j) are large numbers (symbolically - ∞); in the second case, the opposite is true, elements of covariance matrix **P**(i,j) are large numbers, while elements of **I**(i,j) are small numbers. To summarize, when our initial knowledge of system/problem is scarce, we may change our estimation space and move to information space in which this lack of knowledge is easily defined with $\mathbf{I} = \mathbf{0}$. This is the main reason of EIF deployment in this paper; we would like to model and test those problems in which designers

have little initial knowledge of the system. For additional EIF and EKF analysis the reader is referred to [3-5].

B. EIF based training of Hyper Basis Function Neural Networks with Growing and Pruning Ability

In this section we briefly introduce and explain EIF HBF-GP training algorithm; for additional information and deeper understanding of advanced theoretical concepts the reader is referred to [2]. Firstly, we model the input density $p(\mathbf{x})$ with GMM, which is why closed form analytical solution is enabled; now, we may define the concept of neuron significance [2]:

$$\hat{E}_{sig}(k) = \|\mathbf{w}_k\|_q \left((2\pi/q)^{n_x/2} \det(\mathbf{\Sigma}_j)^{-1/2} \mathbf{N}_j^T \mathbf{A} \right)^{1/q} \quad (8)$$

where q is the vector norm, $\mathbf{A}$ denotes vector of mixing coefficients of GMM, i.e. $\mathbf{A} = [\alpha_1, \alpha_2, ..., \alpha_M]^T$, and M Gaussian distributions $\mathbf{N}_j$ are defined as:

$$\begin{aligned} \mathbf{N}_j = \Big[& N(\boldsymbol{\mu}_j - \mathbf{v}_1; \mathbf{0}, \mathbf{\Sigma}_j^{-1}/q + \mathbf{\Sigma}_1), ..., \\ & N(\boldsymbol{\mu}_j - \mathbf{v}_M; \mathbf{0}, \mathbf{\Sigma}_j^{-1}/q + \mathbf{\Sigma}_M) \Big]^T \end{aligned} \quad (9)$$

where $N(\mathbf{v}_k, \mathbf{\Sigma}_k); k = 1, ...M$ denote k-th Gaussian distribution in GMM. The EIF HBF-GP sequential learning algorithm is given in Table I.

HBF-GP sequential learning algorithm requires several input parameters: initial number of HBF neurons (J_0), parameters required for growth criterion $(\varepsilon_{\min}, \varepsilon_{\max})$, threshold for neuron significance $E_{\min}$, overlap of the hyper basis functions κ, and desired learning accuracy $e_{\min}$.

Prior to beginning of the learning process, the estimation of input density $p(\mathbf{x})$ requires $N_{pre\_history}$ data points.

Algorithm continues with calculation of parameter ε_i needed for growth criterion. In the same step, the algorithm calculates current error of the network $\mathbf{e}_i$ and nearest neuron k to the newest input sample $\mathbf{x}_i$. Neuron significance is calculated using (6). The next step of the HBF-GP determines whether new neuron should be added, i.e. step 3.4: the HBF-GP adds new neuron to the hidden layer only if the possible new neuron *J+1* is sufficiently far from existing neurons, i.e. $\|\mathbf{x}_i - \boldsymbol{\mu}_k\|_{\mathbf{\Sigma}_k} > \varepsilon_i$. The second criterion is same as in [13, 14], and insures that significance of possible new neuron $\hat{E}_{sig}(J+1)$ is greater than threshold significance $E_{\min}$. When neuron *J+1* satisfies these conditions it is added to the HBF structure, and learning procedure goes back to step 3 and presents new learning pair $(\mathbf{x}_i, \mathbf{y}_i)$. Otherwise, the HBF-GP will not add new neuron *J+1* to the HBF hidden layer. In this case, the learning procedure continues with update of the nearest neuron *k* (calculated at step 3.2) with EIF. The reader may notice that only parameters of neuron *k* nearest to the input sample $\mathbf{x}_i$ are updated with EIF: greater difference between input vector $\mathbf{x}_i$ and the *j*-th prototype vector $\boldsymbol{\mu}_j$ (*j* = 1,…,*J*) will result in smaller output/activation of the Gaussian basis function.

TABLE I.
SEQUENTIAL LEARNING ALGORITHM FOR HYPER BASIS FUNCTION NEURAL NETWORKS WITH GROWING AND PRUNING ABILITY

input $(J_0, \varepsilon_{\min}, \varepsilon_{\max}, E_{\min}, \kappa, e_{\min})$

1. Estimate input density $p(\mathbf{x})$ with GMM to obtain estimate $\hat{p}(\mathbf{x})$

2. Set initial parameters of HBF network: $\mathbf{w}_J, \boldsymbol{\mu}_J, \mathbf{\Sigma}_J$

2.1 Define EIF state vector

$\boldsymbol{\lambda} = [\mathbf{w}_1^T \ \mathbf{w}_2^T \ \dots \ \mathbf{w}_J^T \ \boldsymbol{\mu}_1^T \ \boldsymbol{\mu}_2^T \ \dots \ \boldsymbol{\mu}_J^T$

$\dots \ \mathbf{\Sigma}_1^1 \ \mathbf{\Sigma}_1^2 \ \dots \ \mathbf{\Sigma}_1^{n_x} \ \dots \ \mathbf{\Sigma}_J^1 \ \dots \ \mathbf{\Sigma}_J^{n_x}]^T$

3. **For** each observation $(\mathbf{x}_i, \mathbf{y}_i)$ do:

3.1 Calculate parameters for growth criterion

$\varepsilon_i = \max[\varepsilon_{\max}\gamma^i, \varepsilon_{\min}], (0 < \gamma < 1)$

Calculate current error of the HBF network

$\mathbf{e}_i = \mathbf{y}_i - \mathbf{f}(\mathbf{x}_i)$

Find neuron *k* nearest to the input sample $\mathbf{x}_i$

$k = \arg\min \|\mathbf{x}_i - \boldsymbol{\mu}_k\|_{\mathbf{\Sigma}_k}$

3.2 Calculate parameters for potential new neuron

$\mathbf{w}_{J+1} = \mathbf{e}_i, \ \boldsymbol{\mu}_{J+1} = \mathbf{x}_i, \ \mathbf{\Sigma}_{J+1} = \kappa \|\mathbf{x}_i - \boldsymbol{\mu}_k\|_{\mathbf{\Sigma}_k} \mathbf{I}$

3.3 Compute significance of newly added neuron $J+1$, i.e. $\hat{E}_{sig}(J+1)$

$\hat{E}_{sig}(J+1) = \|\mathbf{w}_{J+1}\|_q \left((2\pi/q)^{n_x/2} \det(\mathbf{\Sigma}_{J+1})^{-1/2} \mathbf{N}_{J+1}^T \mathbf{A} \right)^{1/q}$

3.4 Update parameters of HBF network

If $\|\mathbf{x}_i - \boldsymbol{\mu}_k\|_{\mathbf{\Sigma}_k} > \varepsilon_i$ and $\hat{E}_{sig}(J+1) > E_{\min}$

allocate new unit with parameters $\mathbf{w}_{J+1}, \boldsymbol{\mu}_{J+1}, \mathbf{\Sigma}_{J+1}$

Else

Update parameters of nearest neuron *k* using EIF

$\hat{\boldsymbol{\lambda}}_{t|t-1} = \hat{\boldsymbol{\lambda}}_{t-1|t-1}; \ \mathbf{I}_{k|k-1} = \left((\mathbf{I}_{t-1|t-1})^{-1} + \mathbf{Q} \right)^{-1}$

$\hat{\mathbf{y}}_t = \mathbf{g}(\hat{\boldsymbol{\lambda}}_{t|t-1}, \mathbf{x}(t))$

$\mathbf{I}_{t|t} = \mathbf{I}_{k|k-1} + \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{H}_t; \ \mathbf{K}_t = (\mathbf{I}_{t|t})^{-1} \mathbf{H}_t^T \mathbf{R}_t^{-1}$

$\hat{\boldsymbol{\lambda}}_{t|t} = \hat{\boldsymbol{\lambda}}_{t|t-1} + \mathbf{K}_t (\mathbf{y}(t) - \hat{\mathbf{y}}_t)$

Compute significance of nearest neuron *k*, *i.e.* $\hat{E}_{sig}(k)$

$\hat{E}_{sig}(k) = \|\mathbf{w}_k\|_q \left((2\pi/q)^{n_x/2} \det(\mathbf{\Sigma}_k)^{-1/2} \mathbf{N}_k^T \mathbf{A} \right)^{1/q}$

If $\hat{E}_{sig}(k) < E_{\min}$

Remove *k*-th neuron

EndIf

EndIf

EndFor

output $(J, \mathbf{w}_j, \boldsymbol{\mu}_j, \mathbf{\Sigma}_j), j = 1, ..., J$

The Jacobian matrix $\mathbf{H}_k = \nabla_{\boldsymbol{\lambda}} \hat{\mathbf{y}}_k^i = [\nabla_{\mathbf{w}} \mathbf{f}(\cdot) \ \ \nabla_{\boldsymbol{\mu}} \mathbf{f}(\cdot) \ \ \nabla_{\sigma} \mathbf{f}(\cdot)]$ will have non-zero elements only for the nearest neuron *k*; for other neurons $(\forall j \notin k)$ the Jacobian will approach to zero or small neglecting value [13, 14]. Finally, EIF update of HBF-GP neuron is performed in step 3.4. After update of parameters of the nearest neuron *k*, the algorithm moves

on to calculate the significance of the nearest neuron $\hat{E}_{sig}(k)$ and checks if it is greater than threshold significance $E_{\min}$; pruning of neuron occurs if its significance $\hat{E}_{sig}(k)$ is smaller than $E_{\min}$. This learning procedure is sequentially applied for all training samples; when final sample *N* is processed, the learning stops and algorithm outputs parameters of optimized HBF network. Outputs of the learning algorithm are parameters of HBF network: total number of HBF processing units (*J*), connecting weights $\mathbf{w}_j$, centers $\boldsymbol{\mu}_j$ and widths $\boldsymbol{\Sigma}_j$ of HBFs $(j=1,...,J)$.

IV. Experimental Results

A. Experimental Setup and Intuition

To fully assess performance of EIF we setup a series of experiments to perform fair comparison between EIF and its dual EKF. Experimental evaluation is performed using two well-known and widely recognized problems for system identification of unknown nonlinear systems: (i) dynamical system with long input delays-in which the behavior of the system is to be predicted using previous system output and controls in HBF neural network, (ii) nonlinear dynamical system-the dynamical system is nonlinear and HBF-GP neural network is to "figure out" the behavior of system using previous systems states and previous controls. EIF HBF-GP neural network is compared to EKF HBF-GP neural network. We used the same initial parameters for both neural networks. These simulated engineering problems should provide fair comparison between two sequential learning algorithms of hyper basis function neural network with growing and pruning ability, and provide assessment of their performance in terms of compactness of network topologies and generalization. All codes for HBF-GP neural network and EIF-based sequential learning (optimization of parameters) are written and run in Matlab 7.12 programming environment; all experiments are conducted on laptop computer with Intel(R) Core™ i5-4200U CPU @ 1.6GHz (2.3GHz) with 6GB of RAM, running on 64-bit Windows 7.

B. Dynamical System with Long Input Delays

Consider the dynamical plant represented with following equation [2]:

$$y_p(t+1)=0.72y_p(t)+0.025y_p(t-1)u(t-1)+... \\ 0.01u^2(t-2)+0.2u(t-3) \quad (10)$$

There are two input values $y_p(t)$ and $u(t)$ fed into HBF-GP neural network to generate desired output value $y_p(t+1)$. The training inputs are uniformly distributed in the [-2,2] interval for about half of the training time and a single sinusoid signal $1.05\sin(\pi t/45)$ for the remaining training time. The training data has 1000 training examples. To verify performance of the generated HBF-GP network and analyze identification results, the testing signal is adopted as:

$$u(t)=\begin{cases}\sin(\pi t/25), & 0<t<250,\\ 1.0, & 250\le t<500\\ -1.0, & 500\le t<750\\ 0.3\sin(\pi t/25)+0.1\sin(\pi t/32) & \\ +0.6\sin(\pi t/10), & 750\le t<1000\end{cases} \quad (11)$$

The testing set consists of 1000 examples. The input into HBF neural network is formed by previous system state and the most recent control, i.e. input is given by the vector $\mathbf{x}=\left[y_p(t)\quad u(t)\right]^T$. Figure 1 shows training and testing sets. As stated, HBF neuron enables local scaling of data, which is especially important for dynamical systems [2, 8]. Therefore, in experiments we partitioned weighting matrix $\boldsymbol{\Sigma}_j$ into two blocks, i.e.

$$\boldsymbol{\Sigma}_j=\kappa\cdot diag\left(\boldsymbol{\Sigma}_{y_p},\boldsymbol{\Sigma}_u\right) \quad (12)$$

where first block $\boldsymbol{\Sigma}_{y_p}$ defines spreads of state variables $y_p(t)$, while the second block $\boldsymbol{\Sigma}_u$ defines spread of controls $u(t)$; scalar $\kappa\in\mathbb{R}^1$ represents initial scale. Initial values of weighting matrix $\boldsymbol{\Sigma}$ are adopted as:

$$\boldsymbol{\Sigma}_{y_p}=\left\{\max\left(y_p(t)\right)-\min\left(y_p(t)\right)\right\}_{t=1}^{1000};\\ \boldsymbol{\Sigma}_u=\left\{\max\left(u(t)\right)-\min\left(u(t)\right)\right\}_{t=1}^{1000} \quad (13)$$

HBF parameters are set as: $\varepsilon_{\max}=5$, $\varepsilon_{\min}=1$, and $\gamma=0.99$, the initial state uncertainty for EKF are: $p_0=0.9, \mathbf{P}_0=p_0\mathbf{I}_\mathbf{P}$, state transition uncertainty $q_0=0.001, \mathbf{Q}=q_0\mathbf{I}_\mathbf{Q}$, and measurement uncertainty $r_0=1, \mathbf{R}=r_0\mathbf{I}_\mathbf{R}$, where $\mathbf{I}_\mathbf{P}$, $\mathbf{I}_\mathbf{Q}$, and $\mathbf{I}_\mathbf{R}$ are the identity matrices of appropriate dimensions (not to be confused with information matrix **I** defined and used in (8), (9) and Table I). On the other hand, the initial state uncertainty for EIF is different. Namely, we simulated situation in which engineer is faced with real situation in which he has no knowledge about the problem or its knowledge is scarce, but the neural model has to be developed, tested and implemented. Therefore, we move to information space.

Figure 1 depicts training set (upper half) and testing set (lower half) as given by equation (11). Figure 2 shows performance of HBF-GP neural network trained with EKF (above) and EIF (below) for testing set. As it may be seen, both sequential learning algorithms are able to learn unknown relationship between long input delayes and output, and generate desired testing response.

Experimental results averaged over 30 independent trials are presented in Table 1; Table 1 shows Root Mean Square Error (RMS), Mean Absolute Error (MAE), and number of processing units, all averaged over 30 independent trials.

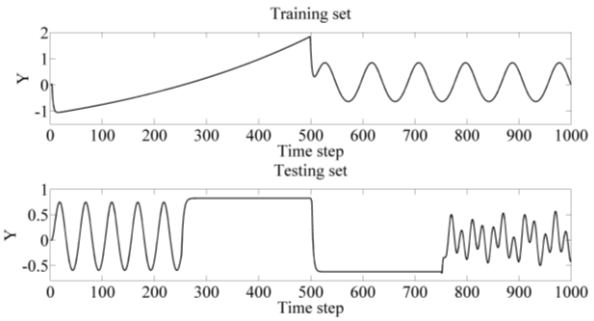


Figure 1. Training and testing set for dynamical system with long input delays.

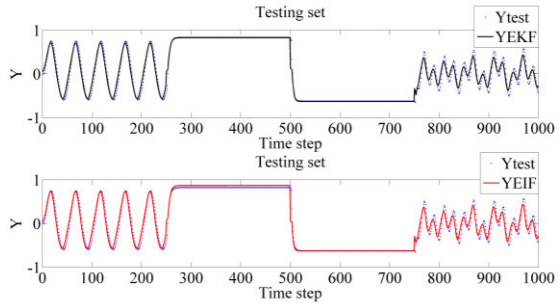


Figure 2. YEKF (EKF HBF – GP) and YEIF (EIF HBF – GP) versus testing set for dynamical system with long input delays.

C. Nonlinear Dynamical System

Consider nonlinear dynamical system given as:

$$y(t+1) = \frac{y(t)y(t-1)y(t-2)u(t-1)\left(y(t-2)-1\right)+u(t)}{1+y^2(t-2)+y^2(t-1)} \quad (14)$$

where $u(t)$ is the control variable (identically independently distributed) in uniform range $[-2,2]$. 800 data points is generated for training of the model. On the other hand, testing set is defined with following dynamics of the control input $u(t)$:

$$u(t) = \begin{cases} \sin(3\pi t/250) & , \quad t \le 500 \\ 0.25\sin(2\pi t/250) + & \\ 0.2\sin(3\pi t/50) & , \quad t > 500 \end{cases} \quad (15)$$

800 data points is used for testing of the model. Figure 3 show training and testing sets. Input vector into HBF neuron is given as $\mathbf{x} = [y(t), y(t-1), y(t-2), u(t), u(t-1)]$, whereas output is given as scalar $y_p = [y(t+1)]$. The initial covariance is initiated using (12) and (13); similarly, we set $p_0 = 0.9, \mathbf{P}_0 = p_0\mathbf{I}_\mathbf{P}$ for EKF while $\mathbf{I} = 0$ for EIF. 30 independent repetitions of HBF-GP learning are conducted. Figure 4 shows how HBF-GP neural network trained with EKF and EIF is able to learn unknown relationship between input vector and output, given training set (14). Experimental results are given in Table II. Furthermore, Figure 5 shows the evolution of total number of HBF neurons during learning process of HBF-GP. Results shown in Figure 5 are averaged over 30 independent trials; as it may be seen, both EKF and EIF converged to approx. three HBF neurons.

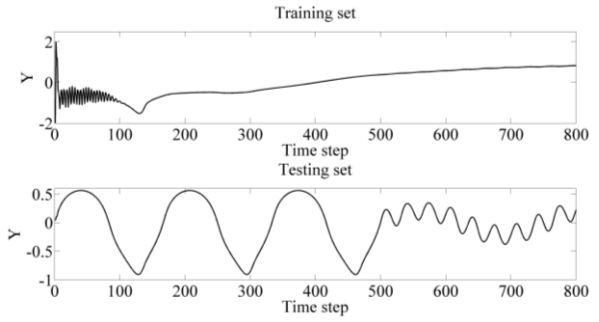


Figure 3. Training and testing set for nonlinear dynamical system.

TABLE I.
EXPERIMENTAL RESULTS FOR HBF-GP NEURAL NETWORK TRAINED WITH EKF AND EIF

	Initial scale of weighting matrix $\mathbf{\Sigma}_j$	Learning algorithm	RMS		MAE		Number of Units
			test	train	test	train	
Dynamical system with long input delays	$\kappa = 0.5$	EKF	0.1025±0.0059	0.0542±0.0069	0.00691±0.0055	0.0388±0.0059	2.9±0.3051
		EIF	0.1141±0.0108	0.0694±0.0112	0.0785±0.0111	0.0524±0.0110	3±0
	$\kappa = 1$	EKF	0.1141±0.0053	0.0630±0.0061	0.0749±0.0046	0.0447±0.0033	2.0667±0.2537
		EIF	0.1258±0.0079	0.0818±0.0067	0.0855±0.0046	0.0587±0.0052	2±0
Nonlinear dynamical system	$\kappa = 0.5$	EKF	0.0524±0.0082	0.1134±0.0182	0.0427±0.0056	0.0554±0.0042	2.9333±0.5833
		EIF	0.0587±0.0174	0.1367±0.0252	0.0477±0.0133	0.0736±0.0131	3.1333±0.6814
	$\kappa = 1$	EKF	0.0717±0.0374	0.1396±0.0213	0.0607±0.0349	0.0705±0.0199	2.6333±0.6149
		EIF	0.0591±0.0377	0.1779±0.020	0.0492±0.0354	0.0836±0.0207	2.1667±0.379

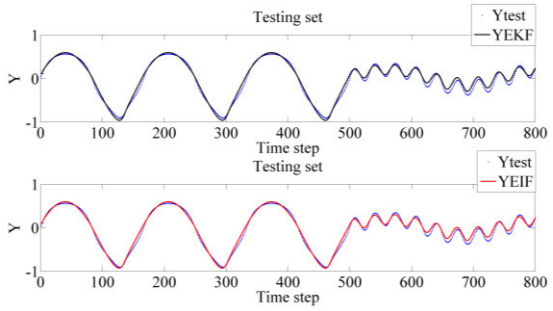


Figure 4. YEKF (EKF HBF – GP) and YEIF (EIF HBF – GP) versus testing set for nonlinear dynamical system.

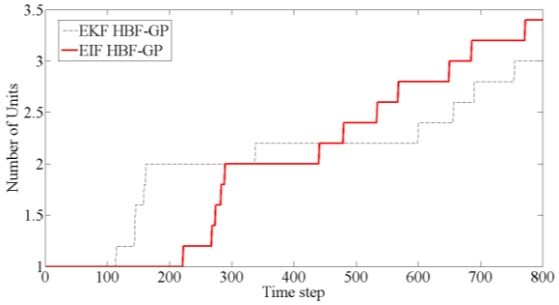


Figure 5. Evolution of number of units during learning for EKF and EIF.

V. Discussion and Conclusion

In this paper, we developed and tested Extended Information Filter (EIF)-based sequential learning algorithm for Hyper Basis Function (HBF) neural network. Unlike a conventional approach, in our research we developed algorithm that enables on line growing and pruning of HBF neural network according to the developed concept of neuron significance [2]. From engineering perspective, EIF has attractive properties when it comes to modeling of complex real world engineering problems with neural networks [3]. Namely, unlike its dual, the extended Kalman Filter (EKF), EIF enables one to set (almost) infinitely large initial covariance matrix; this is important because when faced with hard problems, designer is still able to develop neural network based model/solution although initial knowledge of the problem may be scarce. Furthermore, in our model, we enabled growing and pruning of HBF network topology; the HBF network learns with EIF and simultaneously modifies number of neurons.

EIF is directly compared to its dual (sibling) EKF. As experimental results effectively demonstrate (Figure 1, Figure 4, Figure 5, Table II) EIF-HBF-GP neural network is able to learn complex relations between multidimensional input and output and generate desired response for previously unseen data.

Both of these features are important for engineers working in real world, because real world problems impose mechanisms of how to handle scarce initial knowledge of the problem and how to generate compact network structures. In this paper, we provided a solution for these two problems.

Acknowledgment

This work is supported by the Serbian Government - the Ministry of Education, Science and Technological Development through grant TR35004 (2011-2015).

References

[1] N. Vuković, and Z. Miljković, "Robust Sequential Learning of Feedforward Neural Networks in the Presence of Heavy-Tailed Noise", *Neural Networks*, vol. 63, pp.31-47, April 2015.

[2] N. Vuković, and Z. Miljković, "A Growing and Pruning Sequential Learning Algorithm of Hyper Basis Function Neural Network for Function Approximation", *Neural Networks*, vol. 46C, pp.210-226, October 2013.

[3] N. Vuković, *Machine Learning of Intelligent Mobile Robot Based on Artificial Neural Networks*. Ph.D. dissertation (in Serbian). University of Belgrade – Faculty of Mechanical Engineering, 2012.

[4] N. Vuković, and Z. Miljković, "Machine Learning of Radial Basis Function Neural Networks with Gaussian Processing Units Using Kalman filtering– Introduction", TEHNIKA (in serbian), Vol. LXIX, No. 4, pp. 613-620, 2014.

[5] N. Vuković, and Z. Miljković, "Machine Learning of Radial Basis Function Neural Networks with Gaussian Processing Units Using Kalman filtering – Experimental Results", TEHNIKA (in serbian), Vol. LXIX, No. 4, pp. 621-628, 2014.

[6] N. Vuković, Z. Miljković, B. Babić, and B. Bojović, Training of Radial Basis Function Networks with H^{∞} Filter – Initial Simulation Results, Proceedings of the 6th International Working Conference "Total Quality Management-Advanced and Intelligent Approaches", pp. 163-168, Belgrade, Serbia, 2011.

[7] Z. Miljković, and D. Aleksendrić, *Artificial neural networks—solved examples with theoretical background* (In Serbian). University of Belgrade-Faculty of Mechanical Engineering, Belgrade, 2009.

[8] S.A. Billings, H.L. Wei, and M.A. Balikhin, "Generalized multiscale radial basis function networks", *Neural Networks*, vol. 20(10), pp. 1081-1094, 2007.

[9] T. Poggio, and F. Girosi, "A theory of networks for approximation and learning", A. I. Memo 1140, MIT, 1989.

[10] T. Poggio, and F. Girosi, "Networks for approximation and learning", *Proceedings of the IEEE*, pp. 1481-1497, 1990.

[11] K. Nishida, K. Yamauchi, and T. Omori, "An Online Learning Algorithm with Dimension Selection Using Minimal Hyper Basis Function Networks", *Systems and Computers in Japan*, vol. 37(11), pp. 11-21, 2006.

[12] R.N. Mahdi, and E.C. Rouchka, "Reduced HyperBF Networks: Regularization by Explicit Complexity Reduction and Scaled Rprop Based Training", *IEEE Transactions on Neural Networks*, vol. 22(5), pp. 673-686, 2011.

[13] G.B. Huang, P. Saratchandran, and N. Sundararajan, "An Efficient Sequential Learning Algorithm for Growing and Pruning RBF (GAP-RBF) Networks", *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 34(6), pp. 2284–2292, 2004.

[14] G.B. Huang, P. Saratchandran, and N. Sundararajan, "A generalized growing and pruning RBF (GGAP-RBF) neural network for function approximation", *IEEE Transactions on Neural Networks*, vol. 16(1), pp. 57–67, 2005.

[15] M. Bortman, and M. Aladjem, "A Growing and Pruning Method for Radial Basis Function Networks", *IEEE Transactions on Neural Networks*, vol. 20(6), pp. 1039-1045, 2009.

[16] M. A. T. Figueiredo, and A.K. Jain, "Unsupervised learning of finite mixture models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3), 381-396, 2002.

[17] D. Simon, "Training Radial Basis Function Neural Networks with the Extended Kalman Filter", *Neurocomputing*, vol. 48, pp. 455-475, 2001.

Multi-Objective Tire Design Optimization by Artificial Neural Networks

Miloš Madić*, Nikola Korunović*, Miroslav Trajanović*, Miroslav Radovanović*,
*University of Niš, Faculty of Mechanical Engineering/Department for Production, IT and Management, Niš, Serbia
madic@masfak.ni.ac.rs
nikola.korunovic@masfak.ni.ac.rs
traja@masfak.ni.ac.rs
mirado@masfak.ni.ac.rs

***Abstract*— High performance tire design calls for multi-objective optimization of tire design parameters. This paper discusses the application of artificial neural networks (ANNs) for determination of optimal tire design parameters for simultaneous minimization of strain energy density at belt edge and chafer. Based on finite element (FE) simulation experimental trials, conducted according to full factorial design where three tire design parameters were arranged (belt angle, belt cord spacing and elasticity of tread compound), two ANN models of the same topology were developed. The set of optimal tire design parameter values was obtained by graphical optimization method. The quality of multi-objective optimization solutions was validated by performing additional FE experimental trials.**

I. INTRODUCTION

Although at the first glance it may not be evident, pneumatic tire represents a complex structure, comprising of various rubber components and rubber based composites. Designed for tough exploitation conditions, it must perform well considering a number of mutually opposing performance characteristics such as dry/wet handling and traction, endurance, wear resistance, ride comfort, rolling resistance, aquaplaning, weight, noise and vibration etc. [1, 2]. In order to design a high performance tire that meets, to the greatest extent, desired performance characteristics, selection of suitable tire design parameter values is of prime importance. The main difficulty with which design engineers are faced is the fact that optimal combination of tire design parameter values for one performance characteristic may not even be near optimal for another performance characteristic. From these reasons, for the considered performance characteristics, one needs to formulate and solve tire design multi-objective optimization problem so as to determine suitable combination of tire design parameters.

For tire design optimization different methods and approaches were previously proposed and applied including artificial neural networks (ANNs) [Nakajima et al., 1999], conventional satisficing trade-off method (STOM) [3], multi-objective genetic algorithm (MOGA) and self-organizing map (SOM) [2], utility function approach [Serafinska et al., 2013] and regression analysis (RA) and GA [4]. In most cases, tire design optimization is performed as a two-stage approach: mathematical modeling and optimization. Although the use of RA speeds up and simplifies mathematical modeling process, the use of RA may be of limited applicability and reliability in cases where there exist complex non-linear relationships between dependent and independent variables. As a consequence, the optimization results may not be satisfactory, i.e. there may exist big deviations between experimental and RA model predictions, particularity in the case of multi-objective optimization. In such situations RA polynomial models can be replaced with ANNs, which are based on matrix-vector multiplications combined with nonlinear (activation) functions. Actually, the advantage of the applications of ANNs for empirical modeling of complex non-linearities and interactions in tire design is well documented [2, 5, 6].

Motivated by the lack of studies regarding multi-objective optimization of tire design this paper aims at determination of tire design parameters for multi-objective optimization of strain energy density at belt edge and chafer by the application of ANNs. Determination of the optimal tire design parameter values was performed by graphical optimization method.

II. EXPERIMENTAL PLAN AND FE ANALYSES

As described in detail in [4], objective functions and tire design parameters that were involved in optimization were selected based on two criteria. Those were the significance considering tire design and simple change of tire design parameters inside finite element (FE) model (Fig. 1) used to perform the experiments. Detailed description of the methodology used in finite element modeling and analysis of tires may be found in [7-9].

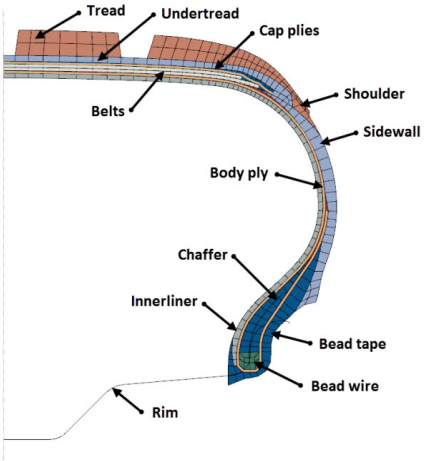


Figure 1. Axisymmetric FE model of an existing tire used to perform the experiments in optimization studies. Structural components of the tire may clearly be distinguished

Three tire design parameters, namely belt angle (A), belt cord spacing (B) and elasticity of tread compound (C) were considered. It should be noted that all tire design parameters are continual variables, i.e. they can take any value within the specified ranges. The tire design parameter ranges were selected based on preliminary FE experimental trials as well as by considering some technically manageable ranges and guidelines from literature. The selected parameters are known to have a significant influence on tire performance, such as maneuverability, durability or rolling resistance. FE experimentation was conducted as per 3^3 full factorial experimental plan upon which each tire design parameter was changed at low, middle and high level. Tire design parameters and their levels within the FE experimentation are given in Table 1.

TABLE I.
TIRE DESIGN PARAMETER RANGES AND THEIR LEVELS

Tire design parameter	Unit	Level		
		1	2	3
Belt angle (A)	º	18	22	26
Belt cord spacing (B)	mm	0.65	1.05	1.45
Elasticity of tread compound (C)		0.6	1	1.4

The configuration of the initial tire design is defined as A = 22º, B = 1.05 mm and C = 1. Tread compound was modeled using hyperelastic Yeoh material model. The value of C = 1 corresponds to nominal values of Yeoh coefficients: C_{10}=1.0236 N/mm^2, C_{20}= -0.4272 N/mm^2 and C_{30}=0.1732 N/mm^2. Values of Yeoh coefficients used in various FE models were obtained by multiplication of all the coefficients with the value of elasticity of tread compound (C). Therefore, elasticity of tread compound (C) is dimensionless.

After conducting 27 FE experimental trials with different combinations of tire design parameter values, values of strain energy density at belt edge (f_1) and chafer (f_2) were recorded and used for development of ANN models. As explained in [4], strain energy density is seen to be a good indicator of complex stress-strain state at a given location inside the tire, taking into account material nonlinearities. Belt edge and bead area are known to be critical zones in tire structure, as abrupt stiffness changes and cyclic flexion lead to stress concentration and fatigue, which in turn cause structural failures.

III. ARTIFICIAL NEURAL NETWORK MODELS

A. ANN Basics

ANNs are one of the most powerful artificial intelligence (AI) tools for mathematical modeling of the relationships between a number of inputs and outputs. Universal function approximation capability, resistance to noisy or missing data, good generalization capability, adaptive nature and other useful features of ANNs made them a preferable choice for modeling complex relationships which are difficult to describe using analytical models.

From many developed types of ANNs, the feed-forward ANNs are among the most used ones, because of their simplicity and ease of implementation. Feed-forward ANNs are composed of a number of simple and highly interconnected processors, i.e. neurons, which are grouped into input, hidden and output layer. Establishment of mathematical relationships between input and outputs is based on input to hidden and hidden to output weights, biases of the hidden and output neurons and the use of transfer (activation) functions in hidden and output layer, which enable non-linear data processing.

B. ANN Models for Optimization of Tire Design

In this study ANN models are aimed at establishing mathematical relationships between inputs, i.e. tire design parameters (belt angle, belt cord spacing and elasticity of tread compound) and outputs, i.e. strain energy density at belt edge (f_1) and strain energy density at chafer (f_2).

FE experimental data, obtained from the full factorial experimental design, were used for development of ANN predictive models. FE experimental data were randomly divided into a data subset for ANN training (22 data) and data subset for testing the prediction accuracy of the developed ANN models (5 data). Given that the number of hidden neurons is dependent on the number of data available for training, two ANN models with four neurons in the hidden layer were designed for the purpose of strain energy density prediction. Since it was assumed that there exist some non-linear relationships between tire design parameters and strain energy density, linear and hyperbolic tangent sigmoid activation functions were used in the output and hidden layer, respectively.

In order to determine near optimal combination of input to hidden and hidden to output weights values and weights of biases of the hidden and output neurons, ANN training resembles a necessary step, which has the predominant influence on the prediction accuracy of developed models. For the purpose of ANN training, Levenberg-Marquardt algorithm was applied due to its fast convergence rate and stability. The ANN training process was monitored via the mean squared error. Fig. 2 shows the variation of mean squared error as a function of the number of iterations.

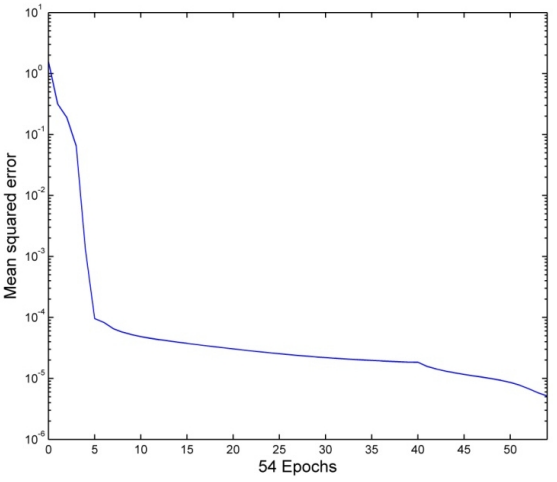


Figure 2. ANN training process

After the ANN training is finished one needs to test the prediction accuracy of the developed models. It is of particular interest to investigate generalization, i.e. ability of ANNs to make accurate predictions when data, which were not used in the training process, are introduced. For the ANN model which related tire design parameters and strain energy density at belt edge (f_1), the mean absolute

percentage errors were found to be 0.03 % and 0.12 % considering training and testing data, respectively. Similarly, for the ANN model which related tire design parameters and strain energy density at chafer (f_2), the mean absolute percentage errors were found to be 0.014 % and 0.034 % considering training and testing data, respectively, which is better then RA modeling as reported in [4]. These statistical results irrefutably confirm excellent agreement between FE experimental data and ANNs predictions as well as high robustness of the developed ANN models. Therefore, these models can be used for the analysis of the effects of tire design parameters on strain energy density as well as to serve as fitness functions for the purpose of tire design optimization.

IV. Analysis And Discussion

The interaction effects of the tire design parameters on the strain energy density at belt edge and chafer are given in Fig. 3. 3-D response surfaces for strain energy density were generated by changing belt angle (A) and elasticity of tread compound (C) at a time, while belt cord spacing (B) was held at low, center and high level.

From Fig. 3 it can be seen that the increase in belt angle (A) results in increase of the strain energy density at belt edge and at chafer. This is probably due to the fact that with increasing belt angle the angle between carcass and belt cord spacing becomes smaller and thus the stiffness change at belt edges becomes larger [4]. It can be also observed that increase in elasticity of tread compound (C) produces a nonlinear increase in strain energy density at belt edge. On the other hand, elasticity of tread compound (C) has negligible influence on the strain energy density at chafer.

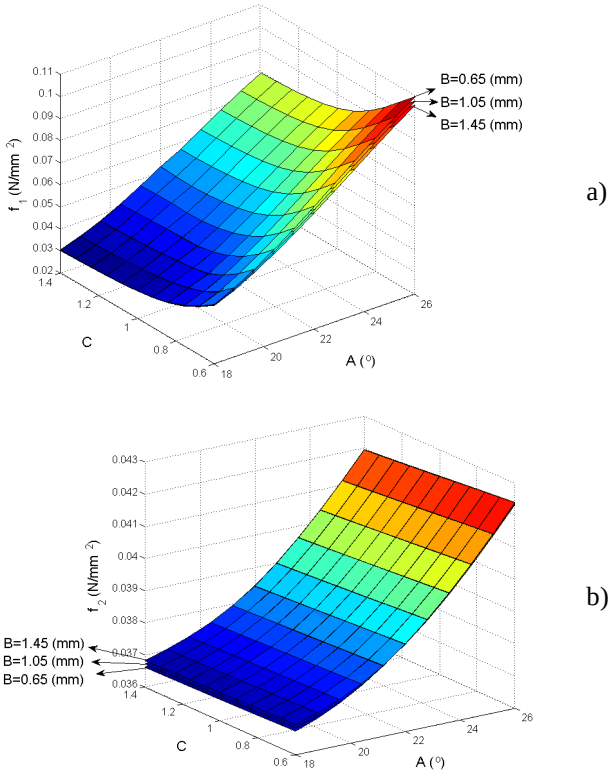


Figure 3. 3-D response surfaces for strain energy density a) at belt edge, b) at chafer

Finally, one can observe that there exists a small decrease in strain energy density at belt edge with increase of belt cord spacing (B). However, regarding the strain energy density at chafer, small decrease in strain energy density comes with decrease of belt cord spacing (B).

From Fig. 3 it is obvious that belt angle (A) has the maximum influence on the strain energy density and that the minimal strain energy density at belt edge and at chafer are obtained when belt angle (A) has minimal value, i.e. A=18º.

The optimal selection of tire design parameters should increase tire durability to some extent by minimizing strain energy density at belt edge and chaffer [4]. Therefore, in the context of multi-objective optimization, the goal is to determine suitable combination of tire design parameters so as to minimize strain energy density at belt edge and chaffer simultaneously. The common approach for multi-objective optimization is based on the use of optimization algorithms. However, based on the conducted analyses multi-objective tire design optimization can be reduced to multi-objective problem having only two independent variables, i.e. belt cord spacing (B) and elasticity of tread compound (C), hence the simplest way for performing multi-objective tire design optimization is graphical optimization method. To this aim, two 3-D response surfaces for strain energy density at belt edge and at chafer are given on the same plot (Fig. 4). Fig. 4 was generated by changing belt cord spacing (B) and elasticity of tread compound (C) at a time, while belt angle (A) was kept constant at A=18º.

From Fig. 4, it is obvious that response surface for strain energy density at chafer is flat, which means that changing belt cord spacing (B) and elasticity of tread compound (C), when belt angle is A=18º, has negligible influence on strain energy density at chafer. Therefore, since the change in strain energy density at chafer is very small, multi-objective tire design parameter optimization problems can be reduced to single objective optimization problem where to goal is to identify tire design parameter values so as to minimize strain energy density at belt edge. The analysis of Fig. 4 reveals that there are different combinations of belt cord spacing (B) and elasticity of tread compound (C) that yield acceptable solutions regarding strain energy density at belt edge. For example, belt cord spacing of B=0.65 mm, elasticity of tread compound C=1.4 produces minimal strain energy density at belt edge of f_1= 0.0299 N/mm^2.

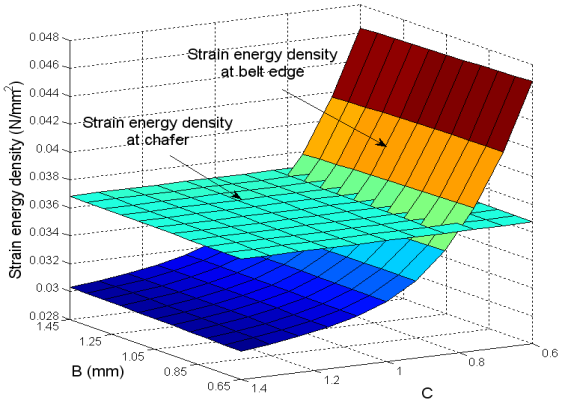


Figure 4. Strain energy density at belt edge and chafer for belt angle A=18º

TABLE II.
COMPARISON OF ANN PREDICTIONS AND FE SIMULATION EXPERIMENTAL VALUES FOR STRAIN ENERGY DENSITY

Optimization solution	A (º)	B (mm)	C	ANN predictions		FE simulation	
				f_1 (N/mm^2)	f_2 (N/mm^2)	f_1 (N/mm^2)	f_2 (N/mm^2)
1	18	0.65	1.4	0.0299	0.0366	0.02997	0.03662
2	18	0.97	1.4	0.0301	0.0367	0.03015	0.03671
3	18	0.73	1.4	0.03	0.0366	0.03002	0.03664
4	18	0.83	1.4	0.03	0.0367	0.03008	0.03667
5	18	0.75	1.4	0.03	0.0366	0.03003	0.03665
6	18	0.65	1.23	0.0306	0.0366	0.03093	0.03664
7	18	0.65	1.36	0.0301	0.0366	0.03021	0.03663

Thus, from the point of view of the performed optimization, any pair (B, C) that yield minimal value of strain energy density at belt edge can be selected as optimal one. In practice, other constructive parameters and goal functions would be taken into account, as well as production limitations and standards, and the choice of parameter values would certainly not be so wide.

In order to check the quality of determined optimization solutions, one needs to compare ANN model predictions and FE simulation experimental values for strain energy density at belt edge and chafer. Thus, several FE simulation experimental trials with the combinations of tire design parameters as given in Table 2 were conducted.

As could be observed from Table 2, there exist a perfect match between values of strain energy density at belt edge and chafer predicted by ANN models and obtained by FE simulation. It can be shown that mean absolute percentage errors regarding strain energy density at belt edge (f_1) and chafer (f_2) are less than 0.5%. These results indicate that ANNs can be efficiently used for multi-objective optimization of tire design parameters.

Regarding initial tire design (A=22 º, B=1.05 mm and C=1) each optimization solution from Table 2 significantly minimizes strain energy density at belt edge (approximately 42.5%) and strain energy density at chafer (approximately 5 %).

V. CONCLUSION

This paper aimed at application of ANNs for determination of tire design parameter values (belt angle, belt cord spacing and elasticity of tread compound) for multi-objective optimization of strain energy density at belt edge and chafer, which are known to influence tire durability. FE simulation based experimental trials, conducted according to full factorial design, provided a set of data for ANNs model development. The conclusions drawn can be summarized by the following points:

- Statistical results indicate excellent agreement between FEM based experimental results and the ANN predictions, which confirms the validity on the use of ANNs for tire design modeling and optimization.
- Quite basic ANN model architecture, trained with Levenberg-Marquardt algorithm using relatively small training data set, outperformed RA based modeling and optimization considering prediction accuracy and generalization capability.
- Belt angle has the most dominant effect on the strain energy density at belt edge and chafer, followed by the elasticity of tread compound and belt cord spacing that have a much smaller influence.

Because of dimension reduction, the optimal tire design parameter values were obtained by graphical optimization method and corresponding strain energy density values were very close to experimentally obtained ones. The determined combinations of tire design parameter values significantly improved initial tire design by simultaneous minimization of strain energy density at belt edge and chafer.

REFERENCES

[1] A. N. Gent and J. D. Walter, *The pneumatic tire*, Washington D.C., National Higway Traffic Safety Administration, U.S. Department of Transportation, 2006.

[2] M. Koishi and Z. Shida, "Multi-objective design problem of tire wear and visualization of its Pareto solutions", *Tire Science and Technology*, 34(3), pp. 170-194, 2006.

[3] J. R. Cho, H. S. Jeong and W. S. Yoo, "Multi-objective optimization of tire carcass contours using a systematic aspiration-level adjustment procedure", *Computational Mechanics*, 29(6), pp. 498-509, 2002.

[4] N. Korunović, M. Madić, M. Trajanović and M. Radovanović, "A procedure for multi-objective optimization of tire design parameters", *International Journal of Industrial Engineering Computations*, DOI: 10.5267/j.ijiec.2014.11.003.

[5] Y. Nakajima, H. Kadowaki, T. Kamegawa and K. Ueno, "Application of a neural network for the optimization of tire design", *Tire Science and Technology*, 27(2), pp. 62-83, 1999.

[6] A. Serafinska, M. Kaliske, C. Zopf and W. Graf, "A multi-objective optimization approach with consideration of fuzzy variables applied to structural tire design", *Computers and Structures*, 116, pp. 7-19, 2013.

[7] N. Korunović, M. Trajanović and M. Stojković, "Finite element model for steady-state rolling tire analysis", *Journal of the Serbian Society for Computational Mechanics*, 1(1), pp. 63-79, 2007.

[8] N.Korunović, M. Trajanović, M. Stojković, D. Mišić and J. Milovanović, "Finite element analysis of a tire steady rolling on the drum and comparison with experiment", *Strojniški Vestnik - Journal of Mechanical Engineering*, 57(12), pp 888-897, 2011.

[9] N. Korunović, M. Stojković, D.Mišić and M. Trajanović, "FEM based parametric design study of the tire profile using dedicated CAD model and translation code", *Facta universitatis, Series: Mechanical Engineering*, 12(3), pp. 209-222, 2014.

REDUCING WAGONS ACCUMULATION TIME IN CLASSIFICATION YARDS BY GENETIC ALGORITHM

Sanjin Milinković*, Rajko Karličić*, Slavko Vesković*, Miloš Ivić*, Ivan Belošević*
* University of Belgrade, Faculty of Transport and Traffic Engineering, Belgrade, Serbia
s.milinkovic@sf.bg.ac.rs, rajkokarlicic@yahoo.com, veskos@sf.bg.ac.rs, m.ivic@sf.bg.ac.rs, i.belosevic@sf.bg.ac.rs

***Abstract* – The process of railcars accumulation is one of the most important operations carried out at marshalling yards as it is the part of train forming process. Accumulation parameter is used for the evaluation as the indicator that measures the efficiency of the processes in the classification yard. It is used to determine the average collection train time per average number of wagons. This paper presents a model for reducing of collection times by genetic algorithm. Model uses genetic algorithm to search for the initial railcar groups from which to start the collection process and thus minimize the accumulation parameter. The model was tested for the Belgrade marshalling yard to calculate the optimal schedule of collection process for direct trains.**

I. INTRODUCTION

Marshalling yards are complex railway stations where freight trains are disassembled and assembled or rearranged in order to create trains according to wagon flows. Trains operations in marshalling yard are planned according to the schedule of the arriving trains. These tactical plans must include details on arriving trains, lists of wagons currently in the yard and other parameters (including type and characteristics of the shunting operation, repairing of the wagons, shunting of part-load shipments and supplying of the refrigerated railcars (wagons) with ice. Trains consist of wagons with different routes of referral, so they can be disassembled in marshalling yard and then assembled to form new trains consisting of wagons with same referral routes. This procedure typically performed in marshalling yards reduces the transport time and increases efficiency.

The operations that are necessary in order to disassemble and assemble trains are performed in receiving yard, in the zone of gravity (hump) and in the classification yard [1]. Classification yard consist of many tracks where wagons from disassembled trains are queued until the collection process is finished. The new train will be formed when the numbers of accumulated wagons reach the predefined criterion for certain direction. The process of accumulation is performed simultaneously with other operations and like all operations the goal is forming trains [2]. In the process of accumulation of wagons it is necessary to take into account all the wagons that are at a given time at the station by routes (tracks in classification yard). The aim of the accumulation is to collect a certain number of wagons and then to form and dispatch the train. The accumulation time is dependent on the number of wagons present in the classification yard, on the average number of wagons per train for each observed route and on the schedule and composition of the arriving trains. Accumulation parameter is mostly used to evaluate the efficiency of the classification yard performance, as well to comparison, and as an indicator of the process of accumulation it depends on the wagons group that has initiated accumulation. Processes in the classification yard are organized according to the limitations from the applied technology, and by the train parameters for the each departing route of referral [1, 3]. The accumulation of wagons by routes is the longest part of the train forming process, so by minimizing the accumulation parameter we are reducing the costs of the transport. The model for reducing the accumulation time should be able to determine the initial group of wagons for each route of referral so that the total collection times in the classification yards are minimal. We have tested the genetic algorithm method for the optimization of the accumulation time.

Genetic algorithms (GA) are evolutionary computation technique based on a similarity with the processes of selection and evolution in nature. Most of the methods that are called genetic algorithms have the following common elements: population of chromosomes, criterion (fitness) function, genetic algorithm operators and random mutation [4].

Genetic algorithms in transport are most often used for solving scheduling and routing problems [5, 6]. In railways problems, genetic algorithms are used for dispatching of train operations, for locomotive assignment problem, for solving railway traffic control conflicts etc. [7, 8, and 9].

The first step in the application of the genetic algorithms method is to represent a defined problem by a string of genes which is called chromosome. After chromosomal representation it is necessary to generate random population and evaluate the fitness of each chromosome in the population. Next step is to create a new population by repeating the application of genetic operators (selection, crossover and mutation). The newly created population is used for further work of the algorithm until a stopping criterion is fulfilled.

Goal of this paper is to examine the possibility of applying genetic algorithms as a decision support model to determine the best solution for initial wagons in the process of accumulation. We have created a hybrid model consisting of GA model (in Matlab [10]) that is connected to the spreadsheet software for input and output data.

II. THE PROBLEM DESCRIPTION

The process of accumulation of the wagons is the process of forming trains [2]. This process is done in marshalling yards. The duration of the process of

accumulation depends on the total number of wagons for a specific referral as well as the number of wagons in the train that will be dispatched. Qualitative factors influencing the process of accumulation wagons are technology in the yard and the pattern of wagon arrivals.

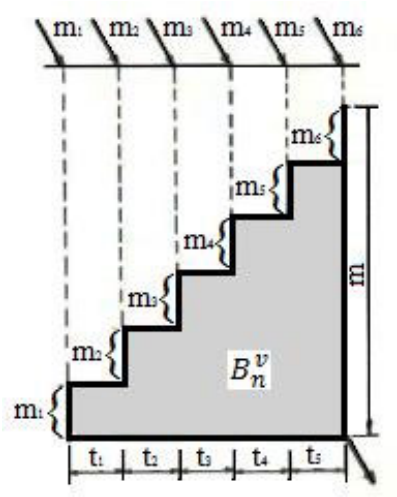


Figure 1. Calculation of train accumulation time

Accumulation parameter (Fig.1) is calculated by:

$$c = \frac{B_n}{m}[h], \text{ for } B_n = \sum t_{gri}\, m_{gri}\ [wagon \cdot h],$$

Where:

B_n – Number of wagon hours of accumulation for the collection of wagons for *i* route of referral,
t_{gr} – Time of group of wagons waiting during the collection process (except for the last, ending group) in hours,
m_{gr} – Number of wagons in a group.

Calculation of the accumulation time (Fig. 1) is by following equation:

$$B_n = m_1(t_1 + t_2 + t_3 + t_4 + t_5) + m_2(t_2 + t_3 + t_4 + t_5) + \ldots + m_5 t_5 + m_0 t_0\ [wagon \cdot h]\ .$$

In order to reduce the time of accumulation it is necessary to perform the following operations: larger groups of wagons should be at the end of accumulation process; and, the largest intervals between group arrivals should be between collected trains.

The process of accumulation should be observed from the moment of wagons arrival at the yard. Operations that are done simultaneously after the arrival are operations at the receiving yard and up to the final operations of forming trains.

Accumulation process is organized by train dispatcher who can prioritize some trains, for example trains in which there are groups of wagons that completes the collection of the train. The complex systems of marshalling yards can be analysed by simulation modelling in order to include all processes [11, 12]. In this model we observe only direct trains, i.e. trains that are formed for departure to the next station and will not change the composition on the route. The procedure for forming these types of trains is called single-stage sorting. Trains formed by a multi-stage sorting method contains wagons that needed to be humped again in order to form the trains that have groups of railcars sorted by the stations along the route [13, 14].

The aim of the model is to mark railcars initial groups for each direction in order to minimize accumulation parameter for the entire station (for all tracks forming direct trains). Model was tested on the Belgrade marshalling yard. To minimize the accumulation parameter it is necessary to determine from which group of wagons the accumulation process starts. The object of accumulation process is to create trains, in such a way that wagons are accumulated for train compositions of individual referral according to the plan of forming. Trains arrive at the station at the time that is defined in MS Excel spreadsheet input data table for each direction of referral. During accumulation process, groups of wagons that are pending consists of a certain number of wagons. Railcars are accumulated to a given number of wagons in the train and when accumulated to sufficient number of wagons for the formation of a train, train is starting the dispatching procedure. Number of wagons in the train varies for direct routes of referral and its value changes depending on the characteristics of the route and tracks. Input data for the model stored in the Excel tables are: arrival times of groups at accumulation process; and the number of wagons in groups. Excel tables are also used to calculate accumulation parameter for the opening sequence of the initial accumulation groups per direction (or track, as the direction is presented by one track). Output data, accumulation parameter for each combination of initial groups of railcars, is calculated from within the Excel thru a series of calculations. For example, accumulation parameter is calculated for opening sequence of initial groups where all twelve direct routes of referral are given by initial groups defined by *i*-th incoming group on *j*-th track. The input data for GA is also defined in Excel file for a chromosome that shows from which groups the accumulation process starts and at the same time indicates the direction of referral of direct trains.

The Matlab genetic algorithms toolbox (*gatool*) [10] were used in order to calculate the minimum accumulation parameter for the entire station. Matlab uses the data defined in *MS Excel* and by using *GA tools* instantly changes the groups from which accumulation begins. The "cooperation" of *MS Excel* and *GA Toolbox* is in exchanging the data where *MS Excel* is used as a fitness function with predefined analytical routines for calculating the accumulation parameter. Model implemented in *GA Toolbox* use data received from *MS Excel* to find minimum accumulation parameter by examining from which group accumulation should begin. This is achieved by importing the results of the fitness function (accumulation parameter *c*) from MS Excel to the GA model. Fitness function is a variable defined by a complex algorithm. Consequently, the value of the fitness function will change when altering the chromosomes or sequence of initial groups by directions. The GA model examines from which group accumulation should begin in order to find minimum accumulation parameter. In the next step, the GA model imports the results of the fitness function from the Excel fitness function as results of the new

sequence generated by the GA model. Thus, the GA model starts the genetic algorithms optimization process by loading the value of fitness function where the accumulation of wagons starts from the *i*-th group for all 12 routes of referral. Then, it changes the value of the chromosome (by changing the number of groups from which accumulation begins) in fitness function and, after the recalculation of the fitness function, the result is imported as a new input data for the GA model. This procedure is performed until the required stopping criterion is fulfilled, i.e. until the minimum value of accumulation parameter is found.

III. IMPLEMENTATION OF THE MODEL

Belgrade marshalling yard (Fig. 2) was used for the implementation of the model as the biggest station in Serbia. Station is semi-automatic, gravitational single-sided marshalling yard with consecutive arrangement of receiving (14 tracks) and classification-dispatching (48 tracks) yards. According to the technology, it is gravitational marshalling yard with double-track hump. Belgrade marshalling yard dispatches direct and pick-up trains, but in this paper we have tested only direct trains or single-stage sorting. Track numbers in classification yard (routes of referral) are labelled by numbers from 1 to 12 and they include directions to: Pozega, Prijepolje, Nis, Presevo, Dimitrovgrad, Subotica and Sid.

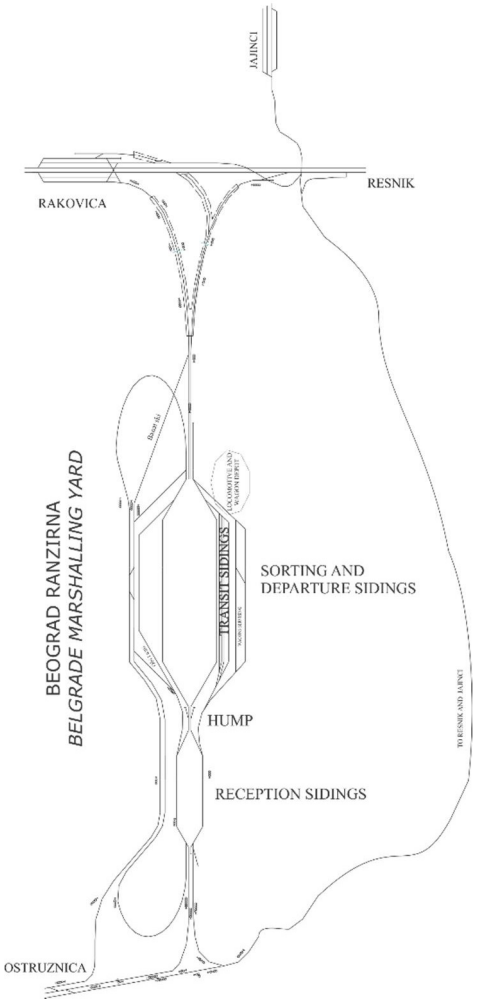


Figure 2. Belgrade Marshalling yard

Input data table is constructed in order to define arrival time of the groups to accumulation process as well as the number of wagons in a group that comes to the accumulation for all of routes of referral. Input data is stochastic and defined by the final reports from the Belgrade marshalling yard. Data on the arrival times, number of railcars, directions, and dispatching times extracted from the official reports are used to define stochastic models. For example, number of wagons in each group varied from one to eight (eight is defined as a limit due to technical limitations of the hump), and uniform distribution was used to define the number of wagons in groups for each route. The initial assumption is that all the processes of the accumulation and then dispatching of trains are finished within 24 hours.

Arrival time of groups at accumulation was generated as a result of two theoretical distributions: Erlang distribution of the second order is commonly used for generating the interval between train arrivals at the station [1], and exponential distribution is used to define the number and directions of wagons in inbound trains. Next step is to calculate the time that wagons spend at the accumulation process. After completion of the process of accumulation it is possible to calculate the parameters of the accumulation of individual and the total accumulation parameter of the entire station. Total accumulation parameter will be dependent on weighted parameters for all 12 routes of referral:

$$c = c_{ni1}\left(\frac{N_{di1}}{\sum N_{di}}\right) + c_{ni2}\left(\frac{N_{di2}}{\sum N_{di}}\right) + \cdots + c_{ni12}\left(\frac{N_{di12}}{\sum N_{di}}\right),$$

where:

c - Accumulation parameter for the station,

$c_{ni1} \ldots c_{ni12}$ - Accumulation parameters by routes of referral,

$N_{di1} \ldots N_{di12}$ - The sum of all wagons coming to the accumulation.

Input data includes the chromosome that is represented as a string consisting of as many digits as there are referral routes.

In order to start the *gatool* in Matlab it is necessary to define the fitness function, variables, the appearance of chromosome (defined in MS Excel), as well as all the other parameters to provide the best solution.

The initial appearance of chromosome can be changed in each generation, depending on time constraints or depending on the objective function. Changes in chromosome will be adjusted in GA Toolbox.

Fitness function c (accumulation parameter for the Belgrade marshalling yard) is defined in Matlab as an external parameter and retrieved from the Excel for each iteration. Excel file calculates the accumulation parameter c for a given input X, where X is defined as string of initial groups for accumulation.

In addition to using Genetic Algorithm Optimization Tool it is possible to run genetic algorithm function directly from a command line. Starting of genetic algorithm function from the command line was performed in the following way:

1. *Xmin* and *Xmax* are defined to determine from which group the accumulation starts. The lower limit for X is

equal to 1, and the upper ranges to the number of groups per routes of referral;
2. After the first step it is defined that *X* takes integer values (*X* represents the group from which the accumulation begins);
3. Next, optimization options structure (*opts*) is created with defined population size, number of generations, items related to the selection, crossover and mutation. A structure options is passed to the optimization function later on;
4. When the optimization options structure is created, genetic algorithm function is initiated;
5. After reaching stopping criterion, final solution is obtained i.e. accumulation parameter of Belgrade marshalling yard.

Variable *X* takes integer values, so there are certain restrictions when setting up optimization options structure. Some of the more important restrictions are:

- Only *doubleVector* population type.
- No custom creation function (*CreationFcn* option), crossover function (*CrossoverFcn* option), mutation function (*MutationFcn* option), or initial scores (*InitialScores* option).
- Genetic algorithm uses only the binary tournament selection function (*SelectionFcn* option), and overrides any other setting.
- No hybrid function. Genetic algorithm overrides any setting of the *HybridFcn* option.
- Genetic algorithm ignores *ParetoFraction*, *InitialPenalty* and *PenaltyFactor* options.

The genetic algorithm attempts to minimize a penalty function, not the fitness function. The penalty function includes a term for infeasibility. This penalty function is combined with binary tournament selection to select individuals for subsequent generations. The penalty function value of a member of a population is:

- If the member is feasible, the penalty function is the fitness function.
- If the member is infeasible, the penalty function is the maximum fitness function among feasible members of the population, plus a sum of the constraint violations of the (infeasible) point.

Given that the fitness function is the same as the penalty function when there is defined area of feasible solutions and in the model bounds for *X* are defined, in the remainder of this paper will continue to be used the term fitness function.

IV. RESULTS

In order to obtain the final solution testing was performed for a population size of 12 individuals, 20 and 24 individuals. These values were chosen based on the recommendation that the population of scale *n* and *2n* is optimal for a specific problem, where *n* is the length of the chromosome [4]. Since we defined the length of chromosome as 12 strings, population size from 12 to 24 was selected. Further, because the testing with a population size of *n* and *2n* only applies to some of the problems, testing was extended for the population size of 50, 100 and 200 individuals. Using a larger number of individuals in the population extends the time searching for solutions but with a larger population it is possible to produce better results.

Along with the change in population size, number of generations was changed and also defined as stopping criterion. Testing was carried out for 24, 50 and 100 generations.

Table I shows results for the case of 12, 20, 24, 50, 100 and 200 individuals in the population and 24, 50 and 100 generations. For a certain values of accumulation parameter the algorithm has interrupted its work. The reason for termination is that average change in the penalty fitness value is less than *options.TolFun* (tolerance on the constraint violation) and constraint violation is less than *options.TolCon* (termination tolerance on the function value). Standard value for *TolFun* and *TolCon* is *1.0000e-6* and should not be lower than 1.0000 e-14. According to the tests the lowest value of the accumulation parameter is 8.8578 hours and it is a case when there are 100 individuals in the population and 50 generations. For the next two cases (100 generations and 100 generations with set *TolFun* and *TolCon* to 1.0000-12), the lowest value of the accumulation parameter is repeated, i.e. the algorithm interrupts operation before completing all 100 generations.

TABLE I. VALUES OF ACCUMULATION PARAMETER FOR DIFFERENT CASES

Population	Generation			
	24	50	100	100*
12	9.59	9.59	9.59	9.59
20	9.71	9.34	9.34	9.34
24	9.21	9.15	9.15	9.15
50	9.20	8.92	8.92	8.92
100	8.96	8.86	8.86	8.86
200	9.33	9.33	9.33	9.33

100* - 100 with *TolFun* and *TolCon*

Fig. 3. and Table I shows changes in fitness function by generations. The number of individuals in a population is 100 and the number of generations 50 and in this case the value of accumulation parameter is 8.8578 hours. This represents the minimal accumulation parameter for Belgrade marshalling yard.

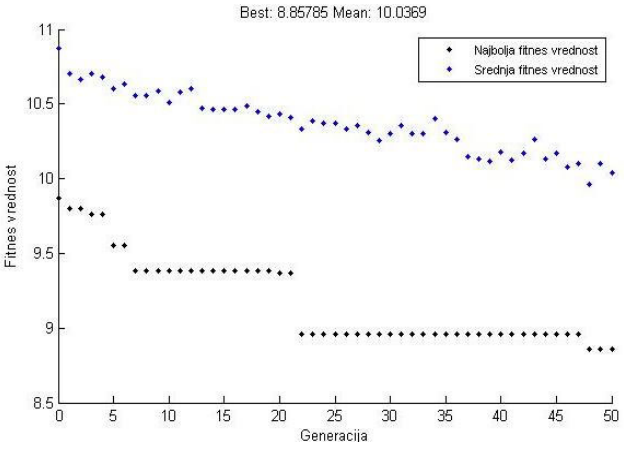


Figure 3. Values of accumulation parameter during GA tool optimization

This value is much lower compared to the initial (when the accumulation process started from the first group of wagons for all directions) which had a value of 9.97 hours. The final value of accumulation parameter was obtained after completion of all stages of the genetic algorithm (generating of initial population, completion of stages of selection, crossover, mutation and fulfilling of stopping criteria). As a stopping criterion the maximum number of generations was used (in this example 50 generations). Appearance of the best solution that is generated in all populations, or at the moment of reaching the final value of the fitness function, is shown in Table II. Namely, the solution obtained by genetic algorithm optimization shows the serial number of incoming group, for each direction, that determines the initial group for train accumulation. Namely, accumulation for the first route of referral (Požega) should start from the 6th group of wagons, for the second route (Prijepolje teretna) from the 3rd group of wagons, for third (Niš and Dimitrovgrad) 5th, for fourth (Preševo) 5th, for fifth (Preševo) 8th, etc.

TABLE II. THE SEQUENCE OF INITIAL GROUPS OF WAGONS FOR ACCUMULATIONON DIRECT ROUTES OF REFERRAL

Route of Refferal	Požega	Prijepolje	Niš	Preševo	Preševo	Dimitrovgrad	Dimitrovgrad	Dimitrovgrad	Subotica	Šid	Šid	Šid
Initial group	6	3	5	5	8	8	11	6	17	5	17	5

In order to check the obtained value of the accumulation parameter, the primary appearance of chromosome was tested for different initial values [15]. This change has produced results of similar error threshold. The best value of accumulation parameter obtained by experimenting with initial value of the chromosome was 8.8578 hours (Table III.). New value of accumulation parameter is reached for the population of 100 individuals and 50 generations. To further verify the final solution graphical representation of the accumulation process was produced (Fig. 4). The accumulation is carried out for each route of referral from the group of wagons that is defined in the final solution. The value of accumulation parameter for Belgrade marshalling yard, which is obtained from graphical method, was 8.8578 hours. This value is the same as in the GA model and thus verifies the solution obtained by GA.

TABLE III. RESULTS FOR THE BELGRADE MARSHALLING YARD

No.	Routes of Refferal (Directions)	Wagons flow N_{di}	Number of Trains N_{vi}	Accumu-lation Wagonhours (B_{nd})	Accumu-lation parameter C_{ni}	$C_{ni}(N_{di}/\Sigma N_{di})$
1	Požega	51	3	131	7.71	0.39
2	Prijepolje teretna	87	5	136	7.53	0.65
3	Niš ranžirna i Dimitrovgrad	74	4	158	8.33	0.61
4	Preševo	66	4	122	7.16	0.47
5	Preševo	70	4	163	9.03	0.62
6	Dimitrovgrad	102	6	154	9.03	0.91
7	Dimitrovgrad	114	6	181	9.51	1.07
8	Dimitrovgrad	76	4	179	9.42	0.71
9	Subotica	76	4	176	9.24	0.69
10	Šid	96	4	241	10.04	0.95
11	Šid	91	4	197	8.56	0.77
12	Šid	110	5	208	9.43	1.02
		1013	53			**8.86**

V. CONCLUSION

The current status of the Belgrade marshalling yard is such that its capacity is not fully exploited. However, in case that the number of freight trains increases, optimization of processes in marshalling yards will be required as the process of accumulation is the most important process regarding the unproductive time for wagons. In technical freight stations the most time is lost for holding up the wagons in the accumulation process. Even with the current number of trains and flow of wagons, with the use of high quality reports on incoming freight trains, optimization of accumulation process could reduce the overall costs of wagons delays at the technical freight stations.

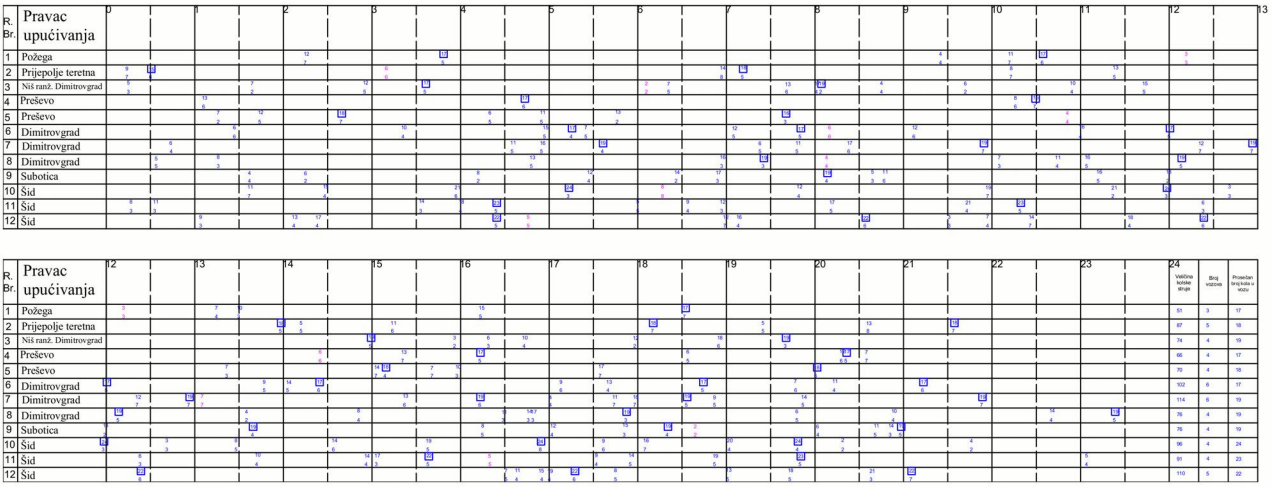


Figure 4. Results obtained by graphical method

Based on the obtained results it can be concluded that the application of genetic algorithms to optimize the accumulation of wagons gives satisfactory results. The obtained results show a significant decrease in the accumulation parameter when compared to initial value that is often in use for creating operational plans. The solution obtained by GA indicates groups of wagons from which the accumulation should begin. As a metaheuristic method, genetic algorithms can produce results that may not represent the optimal solution, but results are close to optimal with small (if any) errors, which is satisfactory for this problem size and structure.

One of the advantages of genetic algorithms is its ability to produce satisfactory results for accumulation parameter that does not require large financial investments. Also, the use of genetic algorithms is not complicated. In further research we will focus on increasing the speed of the model, (i.e. the code that should calculate accumulation parameter within the GA Toolbox) and on the applicability by improving the structure and quality of input data on the arriving trains and wagons. The accuracy of the created model depends on the quality of the input data, i.e. with more precise data on train arrivals and train composition at the station for a certain period of observation, model will be able to produce more precise results.

ACKNOWLEDGMENT

This paper is supported by The Ministry of Educations and Science of the Republic of Serbia, within the research projects No. 36012.

REFERENCES

[1] M. Čičak, S. Vesković, "*Organizacija železničkog saobraćaja II*", Beograd, Srbija, Saobraćajni fakultet, 2005.

[2] N. Boysen, M. Fliedner, F. Jaehn, & E. Pesch, "Shunting yard operations: Theoretical aspects and applications". *European Journal of Operational research*, 220(1), 1-14., 2012.

[3] M. Čičak, "*Modeliranje u železničkom saobraćaju*", Beograd, Srbija, Saobraćajni fakultet, 2003.

[4] M. Mitchell, "*An introduction to genetic algorithms*". MIT press, 1998.

[5] D. Teodorović, M. Šelmić, "*Računarska inteligencija u saobraćaju*", Beograd, Srbija, Saobraćajni fakultet, 2012.

[6] N. Marković, N. Bešinović, and Paul Schonfeld. "Simulation-Based Optimization of Recovery for Multiterminal Freight Transportation System." *Transportation Research Board 91st Annual Meeting*. No. 12-2650. 2012.

[7] S. Dündar, and İ. Şahin., "Train Re-Scheduling with Genetic Algorithms and Artificial Neural Networks for Single-Track Railways". *Transportation Research Part C: Emerging Technologies*, Vol. 27, No. 0, 2013, pp. 1-15.

[8] K. Nachtigall, S. Voget, "A genetic algorithm approach to periodic railway synchronization". *Computers & Operations Research*, 23(5), 453-463., 1996.

[9] P. Tormos, A. Lova, F. Barber, L. Ingolotti, M. Abril, M. Salido, "A genetic algorithm for railway scheduling problems". In *Metaheuristics for Scheduling in Industrial and Manufacturing Applications* (pp. 255-276). Springer Berlin Heidelberg., 2008.

[10] A. J. Chipperfield, and P. J. Fleming. "The MATLAB genetic algorithm toolbox." *Applied Control Techniques Using MATLAB, IEE Colloquium on. IET*, 1995.

[11] P. Marton, N. Adamko, "Villon - a tool for simulation of operation of transportation terminals", *Communication* vol. 10(2)/2008; pp.10-14. 2008.

[12] R. Jacob, P. Márton, J. Maue, M. Nunkesser, "Multistage methods for freight train classification". *Networks*, 57(1), 87-105., 2011.

[13] M. Ivić, A. Marković, S. Milinković, I. Belošević, M. Marković, S. Vesković, N. Pavlović, M. Kosijer, "Simulation model for estimating effects of forming pick-up trains by simultaneous method". In *Proceedings of 7th EUROSIM Congress on Modelling and Simulation*, Prague. 2010.

[14] M. Ivić, I. Belošević, S. Milinković, M. Kosijer, N. Pavlović, "Track properties for formation of pick-up trains". *Građevinar*, 65(02.), 123-134., 2013.

[15] R. Karličić, "*Optimizacija nakupljanja kola primenom genetskog algoritma*". Master rad, Univerzitet u Beogradu – Saobraćajni fakultet, Beograd. 2014.

Simulation model of a Single Track Railway Line

Sanjin Milinković*, Nenad Grubor*, Slavko Vesković*, Milan Marković*, Norbert Pavlović*
* University of Belgrade, Faculty of transport and traffic engineering, Belgrade, Serbia
s.milinkovic@sf.bg.ac.rs, nenad.grubor88@gmail.com, veskos@sf.bg.ac.rs, milan@sf.bg.ac.rs, norbert@sf.bg.ac.rs

Abstract – The analysis of the railway traffic with all complex processes involved can be effectively performed by computer simulation. Simulation modeling is an efficient tool for analyzing railway systems on an operational, tactical and strategic level. We present a discrete model for simulation of railway traffic on a single track railway line. Model uses hierarchy to connect levels and subsystems. State machines and flow charts are used to model how system reacts to events such as train route conflicts. Model was tested on a single track line for different train traffic scenarios.

I. INTRODUCTION

Railway traffic has an important role in mass transport of passengers and goods. Modern railways are in the process of restructuring to be able to compete on the transport market. Railway operators and infrastructure managers are interested in improving the efficiency and utilization of the railway system. This is especially important in the process of timetable design and planning. These processes must be fast and reliable, so there is an increase in computers usage, and specialized software for railway analysis and planning. Software for railway simulation are applied in all levels for operational, tactical and strategic planning.

Important part of planning or reconstruction of rail lines, and for traffic management is an estimation and analysis of the tracks or stations and nodes capacity and utilization [1, 2]. This task is dependent on the complexity and heterogeneity of rail transport and determines how to approach the problem and select the modelling tools. Complexity is correlated with the number of train categories, their speed profiles, and the infrastructure characteristics including number and location of the stations, number of tracks, and with the parameters of the timetables (trains overtaking and conflicting).

The train timetable is a precisely developed plan of train traffic that does not include train delays [3]. Unfortunately, disruptions of train traffic are very common in Serbian Railways. These disruptions can cause delays of trains and affect other trains causing secondary delays. It is not easy to predict disruption processes in a planning process because of theirs stochasticity. For the analysis of the train movements and interactions for a planned timetable it is common to use a simulation modelling approach by applying simulation software that are developed specifically for this purpose [4, 5, 6]. Commercial software packages like *OpenTrack* and *RailSys* [7, 8] use data on trains, infrastructure and timetable information to analyze the train movement and capacity consumption and utilization of the system. It is also possible to define scenarios and investigate impact of the incidents to the train operations and stability of the timetable. These incidents include train failures, infrastructure (signaling and safety) equipment malfunction, humane personnel mistakes and weather and other outside causes (by other transport modes etc.).

The aim of this paper is to present a simulation model developed in *Matlab* [9] that has ability to model a problem: efficient estimation of the single track train operations system parameters. The most important parameters are ones that describe capacity utilization by the number of trains that can operate on the line, utilization of the critical points of the infrastructure (station tracks and switch blocks) and train delays for the defined timetable. Proposed model is able to simulate all the conditions and restrictions imposed by operating rules in Serbian Railways. Further, it implements stochastic disturbances by importing train primary delays (generated by theoretical distribution) into the simulation model. Train primary delays affect the stability of timetable because of their impact to other trains that share same connections or cross their paths generating train conflicts. Because of that, this simulation model has a module for resolving train conflicts. The module based on flow charts theory is located as a subsystem within a simulation model.

II. PROBLEM DESCRIPTION

The simulation model is developed for single track section of a railway line, with 9 stations with defined station tracks and train routes. As an input data for timetable we have defined train categories on the line as well as their speed profiles, acceleration times and times for passenger operations in the stations. Trains are operating in both directions, but some of the local trains are additionally using first two stations.

Following of trains or trains spacing is organized by station section block, that is, there is only one train on a section between two stations [10]. Dwell time for passenger trains is one minute in each station, and stopping time for freight trains is dependent on the current traffic situation. Freight trains have smaller priority then passenger trains, so the dispatching of the freight train is possible only if it does not influence the passenger trains operations. Stations for passing and overtaking are not determined in advance but rather determined during the simulation. Also, it is not possible that freight train can overtake a passenger trains. Freight trains will be operating on the section according to the location and movement of passenger trains, so the dispatching of the freight train will be possible only if the occupation of the next section will not affect the movement of other passenger trains. These rules regulate train traffic with the

higher priority for trains of higher category, i.e. Passenger trains.

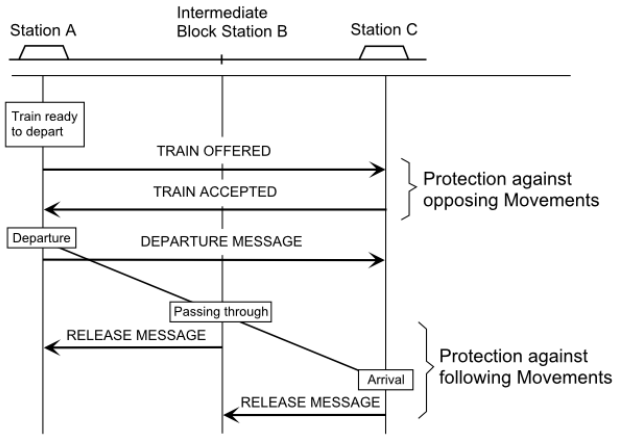


Figure 1. Procedure for train movements on single track line (*source: http://www.joernpachl.de/)*

Simulation model has properties that correspond to typical general decision making logic of a train dispatcher for a single track line. This includes route conflicts situations for same or different priorities (categories) trains. Dispatcher logic is included in the decisions regarding overtaking of trains, passing of trains different categories and spacing of trains. Specific state in the system is a moment when the dispatcher asks for a permission to dispatch train from the station, because received permission includes the information that there is no trains of any priority that will cause a conflict (Fig.1).

III. THE MODEL OF SINGLE TRACK RAILWAY LINE

The simulation model was designed in *Matlab* with tool *Simulink*. *Simulink* tool includes the *SimEvents* tool for simulation of discreet event systems, and tool for modeling of state flow, *Stateflow Chart.* Simulation model is based on discreet events and has a combination of blocks for discreet modeling, blocks for modelling of state flow charts and basic *Simulink* blocks.

Basic concept of the model is a queuing theory, where queues and servers are connected to create subsystems of the train traffic system, with gates to allow or deny a change of state. Open gate is allowing an entity to entry to a next object. Entities represent a train in model, and gates are signals or other conditions that enables the train movement.

Movement of entities thru the model is by subsystems that contain blocks (queues and servers) and decisions are created in the Stateflow Charts subsystem based on the information received from the objects that register the movement of the entities and previous decisions made. Model is organized in two sections (Fig. 2):

- Subsystems that represent station and track sections, and line sections that are connected according to the rail line section plan,
- Diagrams of state flows with implemented decision logic for train traffic management.

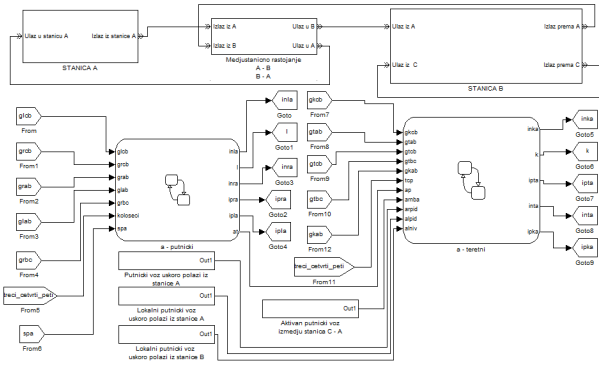


Figure 2. Part of the simulation model

In stations that are used to form a train there is a generation of entities by predetermined schedule of departures (deterministic schedule). After leaving the generating block, entities are assigned with attributes. Attributes are related to parameters regarding train times by sections, train times in stations, and attributes for route choices thru the model (directions of train movements). In starting stations we did not defined the number of track, because there is enough of capacity in those stations. Number of tracks is defined for all of the station along the route on the rail line.

A. Models for intermediate stations

Intermediate stations must be defined so that we can simulate the regulation of train movements. Additionally, we need to define available track capacity and specific use for tracks in each station, for all train categories, for both directions of train movement. We have 7 intermediate stations in the model that are similarly modeled but different by the number of tracks (Fig.3).

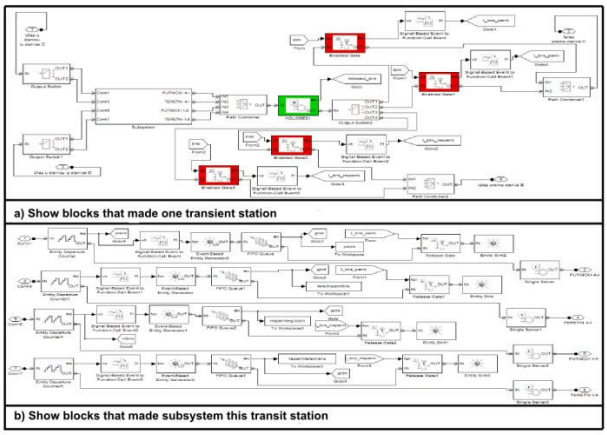


Figure 3. Blocks for representing the intermediate station in the model

Time of train dwelling in stations are determined only for the operations with passengers (leaving and entering the train) but will dependent on the traffic situation including the passing and overtaking of trains and waiting for the clear section (signal to allow the departure). To enable this and record the dwelling or waiting of trains on station track we have used queues on all tracks that have exit signals. Station is modeled to so that all the conditions and rules in real systems are applied to the model. After arriving in the station, entities (trains) are directed to the *OutputSwitch* according to the direction of movement.

From both directions, two categories (passenger and freight) of trains are directed to one of four gates in to the subsystem of the station (Fig. 3a). In the subsystem (Fig. 3b), each entity is directed to the counter (*block Entity Counter)*. Counters have attributes with variables that increase with entry of new entity. Counter information is sent to the block of the signal (Signal Based Event Based to Function Call Generator). Exit port of this block is connected to the block of the entity generator set to generate a new entity by receiving a signal. Entity that has passed the counter generates a new entity that is its copy. A copy is stored in one of four FIFO queues, which represent fictive station tracks. FIFO block, through the port *n*, forwards the information to the *State flow chart* on the number of entities in this queue, and thus provides the knowledge about arriving train's category and direction. The original entity, which has passed through the counter, progresses towards a single channel server with service level from zero minutes to update the attribute value in the counter, and then comes out of this subsystem to be stored on multi-channel server. Multi-channel server specifies a real number of tracks in the station and the arrival of the original entity corresponds to entrance of its copy to the queue. This multi-channel server specifies the service time for entities (i.e. the dwell time of each train in the station). Therefore, all original entities that represent trains are sent to the multi-channel server and their copies are stored inside separate FIFO blocks, according to the train category and direction of movement. The state flow charts gather the total station occupancy by all trains and store information on the category of those trains and the direction of movement.

Multi-channel server is connected with the block for routing entities to multiple outputs. Block for routing entities has four output ports, as two categories of trains can be dispatched to both directions from intermediate stations. Each output port is connected to its block gate (*Gate Enabled*) that enables or disables the departure from the station. If the gate is closed, the output port connected to that gate is blocked and entity cannot enter into the block for routing (*Output Switch*), so entity will remain in the multi-channel server (i.e. on track). According to this feature of block *Output Switch*, to disallow the processing of the entity if its output port is blocked, model can simulate the passing of trains, because one entity can be held in the multi-channel server while other entities are processed through the server. The state flow chart controls working of gates. In Fig. 3a, the first two gates (as viewed from above to the bottom) allow or forbid the departure of passenger and freight train to the right side of station, and the third and fourth allow or forbid the departure to the left side. Thus, it may happen that two gates allow the output to the following combinations: first and third, first and fourth, second and third, second and fourth. These special circumstances correspond to the simultaneous departure of two trains from station on the opposite sides.

After obtaining permission to leave the station, the original entity leaves the multi-channel server, executing the opening of the gate (*Gate Release*) in front of a fictitious queue of copy. After that the copy of original entity is terminated. If the gate *Enabled Gate* is closed for a particular type of entity it means that entity and its copy cannot leave the multi-channel server and the fictional FIFO queue. In this way rule is applied to dwell the entity as is not subject to general rules, such as FIFO or LIFO. Leaving the queue by the original entity depends on the rules implemented in the state flow charts, as they are responsible for the operation of gates *Enabled Gate* that allow or prohibit departure from the station.

This way of modeling intermediate station with embedded adequate rules to control the flow charts enables simulation of trains running in different directions to pass each other or trains heading in the same direction to overtake.

Regarding the sections between neighboring stations, every section between stations is presented using single-channel server. Data about occupancy is also sent to the state flow charts. Upon arrival of entity in single-channel server, its attribute attached during the entity creation could be used to evaluate running time on observed section. This attribute can vary, so trains with different velocities could be observed.

B. Train dispatching rules

Train dispatching is managed by state flow charts and they receive data from blocks in *SimEvents*. There are constructed certain states (*State*), which can be active and inactive. The states are interconnected corresponding to the conditions of transitions (Fig. 4). The states consist of actions that are executed when a given condition becomes active. Conditions are arranged in the term of exclusive *OR* relations, i.e. relation that is mutually exclusive of activity. An active state, means making a single decision, such as the prohibition of new train movement between stations, based on the fulfillment of a condition (occupied open line between these stations).

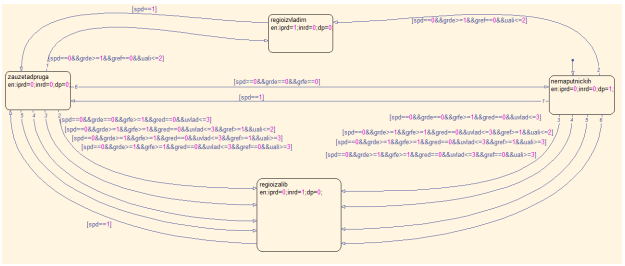


Figure 4. Stateflow chart of the model

Transition assessment from one state to another is performed according to the priorities, so that the execution of one transition means that the other does not execute. Also, it is necessary to define the state to be achieved when the first graph is activated.

State flow charts have the task to regulate the operation of the gates (*Gate Enabled*). Data about the entity position, data about track occupancy in the stations and open line occupancy are sent to the state flow charts. Based on the embedded rules, state flow charts enter into a certain State and define the output data from the charts and forward them to the gates (*Enabled Gate*). These gates should be understood as an output signals in stations or decisions of train dispatcher that permit or prohibit the train movement after he made sure in these decisions (assurance in the model performs state flow charts). Figuratively, this is shown in Figure 5. In the model, an open line between two stations is regulated with two state flow charts so that means that these two charts regulates as much gates as different categories of trains run on this section of the railroad. In addition, one chart is responsible only for the gates through which passing passenger trains

and the other for the gates for freight trains. Hypothetical railroad comprise of 8 open line sections between stations, but in the model there are 16 of these charts.

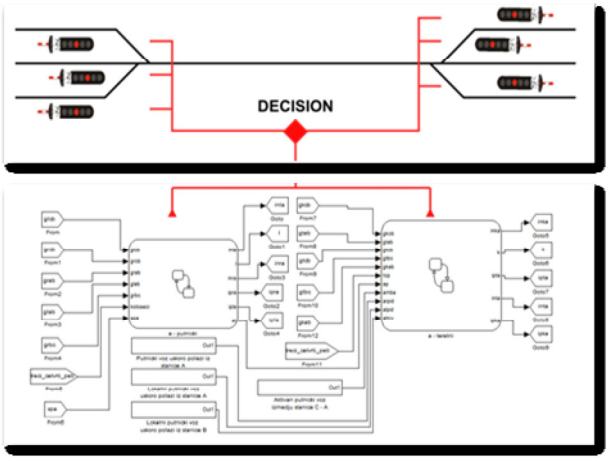


Figure 5. Principle of flow chart modeling

Regulation of entity movement in the model is defined by certain rules. These rules are different for passenger and freight trains, and differ between stations. The rules are set up so that passenger trains have priorities.

In general for this model, the departure of a passenger train from the station is possible, if all the following conditions are fulfilled:

1. The open line between stations should NOT be occupied by other vehicles.
2. There are *n-1* unoccupied tracks in the next station, where *n* is the total number of tracks in the station.
3. There are NO trains with the same rank and the same direction in the next station.
4. If the next station has to depart passenger train in the opposite direction, priority is given to the train which first has met the previous conditions. If both trains have met these conditions simultaneously, priority is given to the train with longer journey travelled.

General rules for the freight train departure from the station implemented in the model include following conditions:

1. Unoccupied distance between stations on which this train should be dispatched,
2. There are less than *n-1* tracks in next station, into which this train is departed, where *n* denotes the total number of tracks in this station,
3. There are no trains of the same rank and the same direction of movement at the departing station,
4. There is no passing passenger train at the departing station,
5. There is no passenger train with the same direction of movement, at the station where this train is located,
6. If there is a passing freight train at the sending train station, priority is given to the train that first meets all previous conditions. If these conditions are met for both trains, priority shall be given to the train that travelled longer journey.
7. If it is estimated that the dispatching of this freight train to the next station will not cause any passenger trains delaying.

The rule No. 7 is complex as it is necessary to combine several parameters to able to process it. Those parameters are collected by inspecting the position of trains on neighboring railway section. This problem is solved by setting up fictitious boundaries of this hypothetically railway section, in which the existence of passenger trains affects the decision of freight trains delaying. Fictitious boundaries are intended to include an area of several train stations and block sections between them. In the model, arrival of the entity that represents a passenger train at a place that represents the beginning of a fictitious boundaries, initiates a copy of that entity who is placed inside of storage block. When the original entity leaves the imaginary boundaries, its copy is terminated. On this way the reservation of the transport route section is completed for the passenger train, so it is unavailable for freight trains until the passenger train left the section. Information about the reservation of this segment is sent to the *Stateflow Chart*, which controls the movement of freight trains within this area. In this way, an absolute priority to the passenger trains is provided.

The set of rules, which can be seen in the Stateflow chart, provide solving the problem of giving permission to the train station, taking into account the train priorities and tracks capacities. It should be noted that during the process of experimenting these rules can be changed depending on the requirement settings.

IV. Model Verification

The duration of the simulation is set to 1440 minutes. We recorded the moments of arrivals, delaying and departures of entities from the subsystems that represent train stations, i.e. tracks in them. Based on these results, the model was verified graphically, by creating the train time diagram, for this hypothetical section of the railroad. The basic layer of train time diagram containing time scale on the horizontal axes and position of stations and length of open line tracks (in suitable drawing scale) is prepared in advance for this case study of single track railway line. Then, using the data obtained from the simulation model (results of the simulation), the diagram is filled with train routes. Events of entity transitions in the model correspond to the moments of appearance of entities in the observed points. These events are points (stations) of the train routes on the diagram. The train time diagram is drawn for the entire period of simulation performance and one part is shown in Figure 6. The *X*-axis represents time of 24 hours and is divided into hours and minutes, where lines of minutes are at 10 minute intervals. The ordinate represents the length of the railway line and intermediate stations are horizontal lines. Blue color shows the routes of freight trains, and black color the routes of passenger trains. In terms of distance, this hypothetical section is designed to suit the route of 70 kilometers.

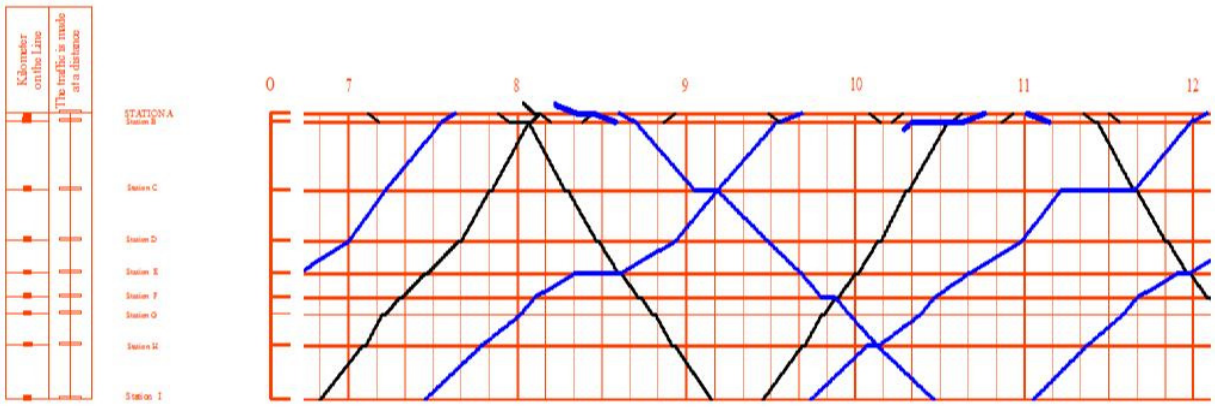


Figure 6. Results presented on a train diagram

V. ANALYSIS OF THE SPECIFIC TRAFFIC SCENARIOS

To better understand results obtained by simulation we have presented them as train time diagrams for two specific scenarios. First scenario is for train overtaking (Fig. 7). During train operations there is a need for overtaking of train when the train with higher speed gain on lower speed train. This is foreseen in the model for the following of two trains of different categories, and it is the case of passenger train overtaking freight train. Figure 7 shows the case of overtaking of slower train, where blue line is a route of the freight train that operates in the model between two stations, and black line shows route of the passenger train.

Train management is implemented by rules imposed by state flow chart. Diagram of the state flow chart that manages the movement of the entity (freight train on Fig.7) had an active state of closed gate until passenger train arrived at the station, and after arrival the state flow changed the transited to the status for allowing freight train further movement. The conditions for the passenger trains were enabling all gates to be opened.

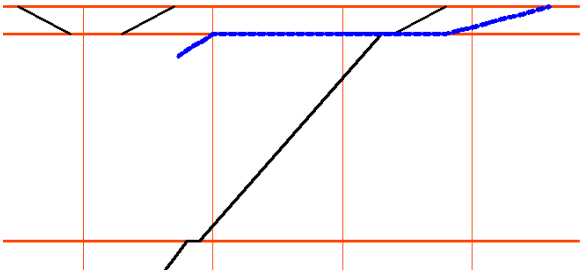


Figure 7. Overtaking of trains in the model

Another specific situation showed on Fig. 8 is for the passing of trains. Passing of trains is event of two train's movement in opposite directions for a single track line. In Fig. 8 freight train (blue line) after arrival at the station is waiting for passing of the passenger train (black line). After the arrival of passenger train to station, the section of the open line is released thus creating conditions for departure of the departure of the freight train.

Figures 7 and 8 shows that the rules and procedures applied in real systems by train dispatchers are successfully implemented in the simulation model, and that for these specific conditions of train traffic model is performing well. Graphical presentation of the results obtained from the simulation enables easier analysis of the modeled system and during the testing assists in verification of the simulation model. Also, graphical results enable the analyst to spot untypical situations and to test possible resolutions for the traffic problems.

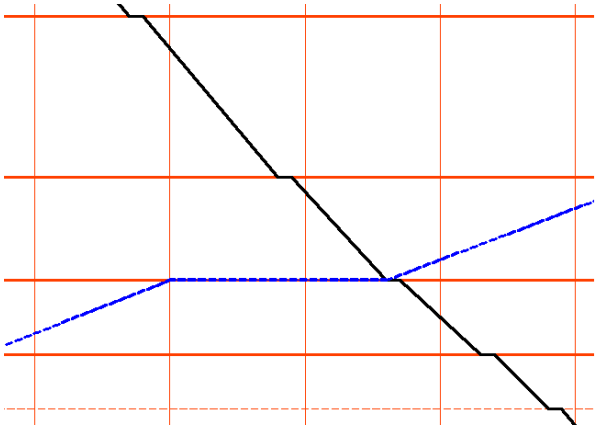


Figure 8. Passing of trains on the singe track line

Results collected in the database and graphical results showed that simulation model is performing according to the set of principles and rules imposed. Detail analysis of the results shows:

- Model has generated the trains as determined in input data with the exact time of departures for each train. Timetable was executed correctly.
- Passenger trains have dwell times in intermediate station as planned by timetable.
- Passenger train did not have delays caused by freight trains movement.
- Model performs adequately according to the rule imposed by the train dispatcher's logic.
- Principles of train spacing applied within a model are verified by graphical results (only one train on one block in one moment of time).

Other results obtained from the model are not presented in this paper as this would require a more detailed presentation on input data, modeling principles and results obtained from experimenting with simulation model. For example, this model was tested with a deterministic timetable, but in testing of the model it is necessary to include stochastic disruptions that cause primary train delays [6].

VI. CONCLUSION

Simulation of the discrete events by Matlab tools enables efficient approach in analyzing of complex systems like single track railway lines. During the development of the simulation model it is possible to include several tools for an integrated approach. Presented model was developed by combining blocks from SimEvents, Stateflow Charts and other basic blocks.

Model for simulation of single track railway line was developed for application in tactical and strategic planning in railways that have operating principles similar to Serbian Railway. Advantage of this model is its modular approach where subsystems can be managed and connected in the same level or form of the hierarchy. This supports building of the simulation models of different level of details, ranging from the micro models to macro models. Another important advantage is the Matlab environment that can be used not just to simulate, but also to include simulation model (as subsystem, or its results) into an optimization model. The disadvantages of the model are its complexity and necessary amount of knowledge and time to learn how to adjust and build the model. Further work on the simulation model will be directed to improve its ease of use and application in the real systems modeling. Also, research is planned for analysis of the data necessary to calibrate the model and for statistical analysis of the disruptions and causes that generates train delays. Another line of research will be directed in developing the Stateflow Chart model for the use in resolving train route conflicts in single track and double track lines and railway stations. Properties of Stateflow Chart such as management of timed events enables the new approach and new applications in modelling complex railway systems.

ACKNOWLEDGMENT

This paper is supported by The Ministry of Educations and Science of the Republic of Serbia, within the research projects No. 36012.

REFERENCES

[1] M. Čičak, S. Vesković, "*Organizacija železničkog saobraćaja II*", Beograd, Srbija, Saobraćajni fakultet, 2005.

[2] M. Čičak, "*Modeliranje u železničkom saobraćaju*", Beograd, Srbija, Saobraćajni fakultet, 2003.

[3] I.A. Hansen (ed.). "*Railway timetable & traffic: analysis, modelling, simulation"*. Eurailpress, 2008.

[4] N. Bešinović, S. Vesković, M. Ivić, S. Milinković, Simulacioni model za utrvđivanje propusne moći pruge Novi Beograd - Batajnica primenom metode UIC 406, in: YU INFO 2011. (Informaciono društvo Srbije, Kopaonik, Srbija), 2011.

[5] N. Grubor, S. Milinković, S. Vesković, P. Márton, "Simulation analysis of the regional railways in South Banat region", *Railway Transport and Logistic*, vol 2013/3, 2013.

[6] S. Milinković, M. Marković, S. Vesković, M. Ivić, N. Pavlović, A fuzzy Petri net model to estimate train delays, *Simulation Modelling Practice and Theory*, 33, 144-157., 2013.

[7] Huerlimann, D. and Nash, A. "Open Track Simulation of railway network Version 1.3.", Institute for Transport Planning and Systems, Zurich, 2010.

[8] U. Fischer, S. Mirković, S. Milinković, A. Schöbel, "Possibilities for integrated timetables within the Serbian railway network", *Facta universitatis-series: Mechanical Engineering*, 10 (2012) 145-156., 2012.

[9] H. Klee, R. Allen. "*Simulation of dynamic systems with MATLAB and Simulink*". Boca Raton, FL: CRC press, 2007.

[10] J. Pachl, "*Railway operation and control"*. EuRailPress. 2002.

Open Satellite Data for the area of Serbia

Dušan Jovanović*, Miro Govedarica*, Filip Sabo*, Dubravka Sladić*
* Faculty of Technical Sciences/Department for Computing and Control Engineering, Novi Sad, Serbia
dusanbuk@uns.ac.rs, miro@uns.ac.rs, filipsabo@uns.ac.rs, dudab@uns.ac.rs

***Abstract*— This paper aims to introduce open access satellite data for the area of Serbia. We describe different satellite platforms which collect images with different spatial and spectral resolution, access and practical utility of historical data and the data that are being collected today. Radar platforms with different bands are introduced as well. We also discuss applicability of these data in different governmental institutions. The paper also indentifies the need to semantically describe such data and proposes introduction of semantic metadata.**

I. INTRODUCTION

The Global Open Data Initiative (GODI), The Open Data Foundation (ODaF), Open Gov Partnership (OGP) are just some of the organizations which define what open data are and how should they be treated by governmental institutions and non-governmental institutions. The definition of open data "Open data and content can be freely used, modified, and shared by anyone for any purpose" [1] is the most common and the most frequently used definition of open data. The Government of the Republic of Serbia in its action plan for OGP implementation initiative for 2014 and 2015 year [2] developed a strategy with 13 planned measures. In the part related to information access, measure 10 refers to new technology introduction to improve services provided to citizens. In that measure, the plan describes necessary actions which would facilitate the access to open data for common citizens. Although, these measures are primarily related to administration authorities, they can refer to spatial data.

The importance of spatial data in the 21 century is increasing exponentially, mainly because of growing availability and usability in everyday life, and because of the vast data quantity which are being collected, processed and analyzed in different ways and by different profit or non-profit institutions. The objective of this paper is to introduce open access remote sensing satellite data for the territory of Serbia which can be good starting point for all those who are in position to use and to analyze remote sensing data.

Usage of this satellite images can significantly improve the efficiency of state organizations dealing with spatial planning, agriculture, environmental protection, waters, floods and other areas in which remote sensing data can be used. These data can be used in educative purposes, in high school and faculty education. Satellite data importance and their open availability can have a significant influence on the mindset of future generations that are now being educated.

This paper is organized as follows: the next section presents what is open access satellite data, what is current international and domestic strategy, and also focuses on historical, current and future remote sensing data. Section three gives an explanation about data applicability in Serbia, and how open access data can be used. Also this section describes the necessity of educating users of remotely sensed data. Next section explains the semantic metadata of satellite images. Paper concludes with an outlook on new technologies in the future.

II. OPEN ACCESS SATELLITE DATA

A. International and domestic strategy

Land cover monitoring, land cover change detection, large cities expansion and their affect on environment, weather disasters, floods, fires, and other occurrences, are all informations that are being monitored and collected on a daily basis. These informations present satellite images with different spatial, spectral, temporal and radiometric resolutions. Property and access to this informations present an important step towards understanding the environment in which we live in and how do we influence on our environment.

Landsat program for Earth observation lasts from 1972 when the first Landsat was launched. Since 1982 Landsat 4 satellite platform began to deliver satellite images with 30 meters spatial resolution in visible, near-infrared and short-wave-infrared wavelengths, which enabled first severe Earth land cover monitoring [3]. A big step forward in information access was in 2008 when the US Geological Survey issued a decision that enabled all of the Landsat images free and open access to anyone [4]. They continued this practice with the launch of new platform Landsat 8 which monitors the Earth with spatial resolution from 15 to 100 meters in visible and infrared (including thermal) wavelengths. Landsat 8 from its launch and until the moment of writing this paper has already made over 450000 images of Earth and all of them are open access. The benefits of open access demonstrated by the Landsat program justify and encourage efforts for much more open access satellite data. Governments and the remote-sensing community should now seize the opportunity to develop a unified strategy for land monitoring [4].

European Space Agency (ESA) within the Copernicus program is constantly developing and planning new Sentinel satellite platforms. On the April 2014 year, the first Sentinel-1 was launched, and more of them are planned to be launched. Sentinel-2 and Sentinel-3 are planned to be launched in 2015. Apart from the different properties of these satellite platforms, common to all of them is the fact that the images that they deliver will be free open access.

Satellite platforms launch into space is not a rarity any more. In the 70's there were only 2 launches by decade, and today that number is more than 10 platforms in one year. Thanks to Google Earth and Bing maps, we now

have more and more people who observe the Earth surface. Within the book that Committee on Earth Observation Satellite's (CEOS) published it can be found that "CEOS agencies are operating or planning around 260 satellites with an Earth observation mission over the next 15 years. These satellites will carry around 400 different instruments" [5].

In [6] the authors identified 197 satellite platforms which were used for the analysis of Earth surface until 2014 year. Furthermore, authors [6] propose the classification of satellite images based on spatial resolution on 5 classes: 0.5-4.9 m (very high resolution), 5-9.9 m (high resolution), 10-39.9 m (medium resolution), 40-249.9 m (moderate resolution) and 250 m-1.5 km (low resolution) and they point out that the images with medium and moderate spatial resolution are generally free through 'free and open' data policies. Images with high and very high spatial resolution are commercial.

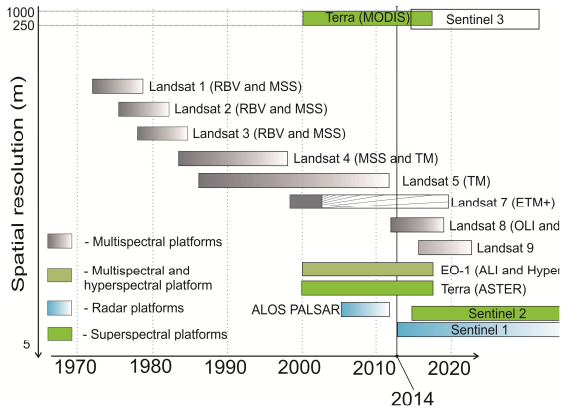


Figure 1. Historical, current and future open access satellite data

If we discuss about satellite images and platforms, Serbia's position does not differ from other countries with similar size and population. Serbia does not possess any satellite platform and does not share the possession of any platform with other countries. If we look at similar countries, members of EU, we can conclude the lack of Serbia's involvement in important EU institutions such as ESA and consequently the lack of involvement in projects such is ongoing Copernicus. Participation in this institutions and programs would certainly have a significant effect on consciousness development regarding the use of satellite images and surely an effect on awareness of open access data. Serbia in its action plan for OGP implementation initiative for 2014 and 2015 year [2] did not predict any measures that would refer to geospatial data access generally, and specially in the light of satellite images open availability. The Republican Geodetic Institution (RGI) as the responsible institution, opened the access to some data that are in their possession (data regarding real estate cadastre and similar) but did not open the access to data that are related to Earth surface on the territory of Serbia. This, remote sensing data refers to satellite images that RGI purchases for own needs, and also to aerial imagery which is used by RGI for different products (ortophotos, terrain models and similar).

B. Historical Data

Figure 1 Historical, current and future open access satellite data, shows all satellite platforms which imaged and still image the territory of Serbia. Spatial resolution of these platforms range from 5 to 1000 meters. Figure 1 also depicts that the platforms are of different spectral characteristics, that is, there are multispectral, hyperspectral and radar images that cover the whole territory of Serbia. Different, open access data with proper systematic treatment, analysis and interpretation would surely help in creating a new, different perspective on the state of Serbia itself.

If we inspect different historical multispectral images which include visible and infrared parts of the electromagnetic spectrum, we can conclude that they present a meaningful information resource for historical land cover reconstruction. Land cover can include: urban area, water area, forests, agricultural land, etc. That is, we can monitor deforestation, floods, expansion of Belgrade, Novi Sad, Nis, and also we can monitor the health of crops, we can simulate crop yield, pollution of water areas can also be monitored, snow cover, determination of soil moisture with active sensors, monitoring the environment and more. Satellite which monitored Earth the most and therefore photographed the territory of Serbia more than every other satellite was Landsat 5. Launched in March 1984, and decommissioned in June 2013, Landsat 5 entered the Guinness book of records as the longest operating Earth observing satellite mission in history. During its mission Landsat 5 collected more than 2.5 million images [7]. Landsat 4 stopped to collect images in December 1993 but was decommissioned in 2001. Combination of multispectral images from Landsat 5 and from Landsat 4 provide a powerful open access archive which can be used together with Landsat 7 and Landsat 8 images in order to monitor Serbia landscape behavior dating from 1982 until now.

Besides historical multispectral images, historical radar images (ALOS PALSAR) are also available for the territory of Serbia for the years: 2007,2008,2009 and 2010. These data are also open access and free. The data include forest and non-forest products as well as the original radar images with two polarizations. The images can be used to monitor forests for mentioned years and for the whole territory of Serbia. Moreover this data can be combined with other images not necessarily radar in order to track changes. The creators of ALOS PALSAR images recommend that images be used for forest monitoring but of course this is not necessary and we can use original radar images to monitor, for example water. The historical data present a significant factor in understanding the ecological changes on Earth during the past decades.

TABLE I.
HISTORICAL DECOMMISSIONED SENSORS WHICH OBSERVED SERBIA

Sensor	Decommissioned	Where to find images
ALOS PALSAR	May 2011	http://www.eorc.jaxa.jp/ALOS/en
Landsat 4	June 2001	http://glovis.usgs.gov http://earthexplorer.usgs.gov
Landsat 5	June 2013	http://glovis.usgs.gov http://earthexplorer.usgs.gov

Table 1 shows the dates when the sensors which cover territory of Serbia have begun to collect images and also the locations on the internet where these images can be found and ordered free of charge. It is necessary only to register as a user and then the images can be downloaded.

C. Current Data

Figure 1 also shows that in the moment of writing this paper, there are 6 satellite platforms which currently observe Serbia and provide open access, that is they are still active and will be. Active platforms image the Earth in visible, near-infrared, middle-infrared, thermal infrared and in the microwave part (Sentinel-1) of the electromagnetic spectrum. Main difference in comparison with historical images is in spatial resolution, which is much higher, with trend towards medium and higher spatial resolutions if the platforms are launched in the last

TABLE II.
SENSORS WHICH MONITOR TERRITORY OF SERBIA

Sensor	Collecting since	Where to find images
ASTER	February 2000	http://glovis.usgs.gov http://reverb.echo.nasa.gov/reverb
EO-1 ALI, Hyperion	November 2000	http://glovis.usgs.gov http://earthexplorer.usgs.gov
Landsat 7	April 1999	http://glovis.usgs.gov http://earthexplorer.usgs.gov
Landsat 8	May 2013	http://glovis.usgs.gov http://earthexplorer.usgs.gov
MODIS	December 1999, May 2002	http://glovis.usgs.gov http://reverb.echo.nasa.gov/reverb http://nsidc.org/data/modis
Sentinel-1	April 2014	https://sentinel.esa.int

few years. This fact, is maybe the best depiction of technology development in the last 30 years, because it was unimaginable that the images with 5 m spatial resolution can be downloaded free of charge.

Table 2 displays the date when the platforms became active and where to find images for territory of Serbia from these platforms. As mentioned before, it is only required to register on the internet and the images can be downloaded.

If the tables 1 and 2 are closely analyzed, it can be seen that all of these datasets, except data from 2 satellite platforms, are controlled by the United States Geological Survey (USGS). Also, all satellite platforms except ALOS and Sentinel-1 are in US property. This information leads to a conclusion that the USGS as one of the most important institution in field of remote sensing, has recognized the significance of making images free and open access to all interested institutions and people. Information that more than 1 million images were downloaded since the Landsat images became free in just one year, whereby the number of bought images were 25000 yearly, speaks in favor of importance of open data access. The second interesting information is that until March 2008 only 7.7% of all Landsat 7 scenes which are stored in the archive were ordered/purchased. After the USGS decision in 2008 to make Landsat images free and open access this percent changed to 65.1% by December of the same year 2008 [8].

Newest satellite platform with open data, Sentinel-1, property of ESA, presents a new breakthrough in the era of free satellite images which EU proclaimed with Copernicus program. Copernicus is the most ambitious Earth observation program to date. It will provide accurate, timely and easily accessible information to improve the management of the environment, understand and mitigate the effects of climate change and ensure civil security.

D. Future Data

Landsat 7 fuel-based end of life is 2017, Landsat 8 fuel ending life is until 2023. Plans for Landsat 9 are still developing, he is planned for December 2018, but the resources from the Congress are not approved yet, neither the goals for the new platform are defined [8].

The Sentinel mission which is conducted by European Commission (EC) in cooperation with ESA, should launch 5 more platforms in order to complete the Copernicus program. First one of 5 is Sentinel-2 with its 13 spectral bands and spatial resolutions which range from 10 to 30 meters. Sentinel-2 is indented to monitor changes in climatic conditions, land cover and in special cases to help in critical situations. If we look at technical characteristics, this platform will be very similar to Landsat platforms under USGS property. The launch is planned for May 2015. Sentinel-3, which will serve in mapping sea surface topography and will monitor sea and land surface temperature with spatial resolutions from 300 to 1000 meters, will be launched in mid of 2015 year, that is the first platform (3A) of 3 will be first launched. The second one, 3B is planned for 2017 and the last one 3C will not be launched before 2020.

We conclude that the Copernicus program is somehow a response from EC to US Landsat program and even more than that. It will be shown in the near future how more.

III. DATA APPLICABILITY

Figure 2 shows some of the results that can be obtained from the open access satellite data for the area of Republic of Serbia. Landsat satellite images can be used for generation of land cover maps, or maps of different vegetation indices that can be used in area of agriculture (quality of agricultural cultivated plants or condition of other types of crops and fruit fields) or in environmental applications. Also, Landsat satellite images can be used

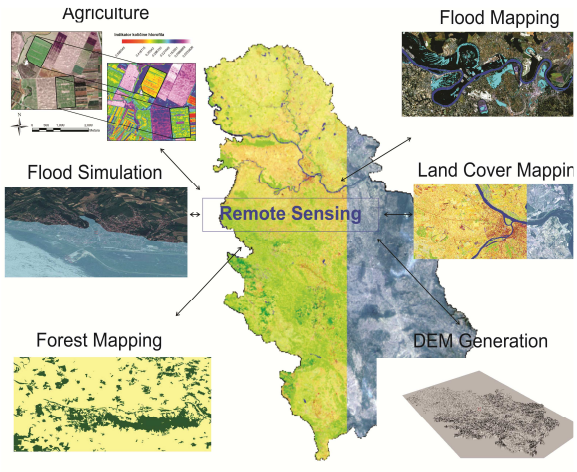


Figure 2. Examples of products obtained from the processing of open access satellite data for the area of Republic of Serbia

for mapping of flooded areas and flood simulations. Figure 2 also illustrates forest map for the area of Fruška Gora National Park, obtained from radar PALSAR images, while Aster images, among other applications, can be also useful for Digital Elevation Model (DEM) generation.

Free satellite images trend lead to increased number of of public and scientific communities which are involved in researching, analyzing and distributing satellite images and products derived from images. This increasing number of users applies to all countries as well as Serbia. This means that more and more people are somehow involved in spatial data processing. Therefore, the necessity of remote sensing education is a must. People need to understand image processing and their potential. Tomorrow these people will work in governmental or public institutions and knowledge of spatial data will surely have a benefit. We are now witnesses of Copernicus ambitious program which will have six platforms in space until 2023 and these data will be free to process and to gain the most from images.

Historical remote sensing images can be used in various applications, in the next paragraph we will mention some of the areas where satellite data for Serbia can be useful.

ALOS PALSAR radar images can be used to map forests for previous years (Figure 2). Mapping forests allows us to monitor deforestation. The effects of vast fires can also be mapped. Also, radar images (Sentinel-1) can be very useful in monitoring water or floods or for DEM generation. That is, today we have flood, and shortly after that we can have maps of flooded areas on the whole territory of Serbia. Historical multispectral images can also monitor any kind of land cover. Popular parameter that can be derived from thermal bands is land surface temperature (LST). Therefore, we can monitor the temperature on any part of Serbia now and in the past. Application in agriculture is constantly increasing. Crop status can be analyzed with infrared bands, soil moisture also, and soil drought. MODIS has special bands that effectively monitor floods and fires, with spatial resolutions from 250 to 1000 m. We hope that Sentinel-3 will also serve in similar purposes. Natural disasters in Serbia are unfortunately increasing and therefore it is necessary to increase the importance of spatial data which can help in monitoring, mapping and better overcoming these disasters.

Authors of this paper can confirm that the bachelor and master theses from remote sensing are very popular among students, and the number of theses is increasing constantly. With new platforms that offer open access, this number will increase even more. That is, people are very interested to work with satellite images. This interest can eventually represent Serbia as a recognized country in satellite image processing.

IV. Open Data Metadata

Effective use of spatial data requires easy access to their documents, or metadata that describes lineage, ownership, quality, maintenance, etc. Considering many different sources of open data it is necessary to describe that data in standardized way using available metadata standards, and make it accessible through metadata catalogues. In this paper we analyze standards that are suitable for describing open geospatial data and propose shift toward the semantic level of metadata in order to make open data more accessible to non expert users.

Currently, the most significant metadata standard for geospatial data is ISO 19115 (ISO 19115-2 is its extension for describing imagery and gridded data) [9]. ISO 19115 classifies metadata standards into categories which include: identification information for data identification, constraint information about access rights, data quality information, information about maintenance of data, spatial representation information, reference system information, information about the content of the dataset and a feature catalogue, distribution information etc. Part 2 of ISO 19115 also introduces imagery identification, such as acquisition information, instrument identification, platform identification, mission identification, processing information etc.

```
MD_Metadata
...
+spatialRepresentationInfo (digital
representation of spatial data)
        MD_Georectefied (descendent of
        MD_GridSpatialRepresentation)
        numberOfDimensions: 2
        axisDimensionsProperties:
                MD_Dimension
                dimensionName: 001 (code for
the ordinate y)
                dimensionSize: 9449 (cell-
pixel)
                resolution:
                        Measure (documented in
ISO 19103)
                dimensionName: 002 (code for
the abscissa x)
                dimensionSize: 14173
                resolution:
                        Measure (documented in
ISO 19103)
        cellGeometry:002 (cell represent
area)
        transformationParameterAvailability:
        1 (transformation parameters are
        available)
        checkPointAvailability: 0 (check
        points for testing accuracy of
        georeferenced data)
        cornerPoints: (geographical
        coordinates of the corners of an
        image)
                GM_Point (documented in ISO
        19107)
        pointInPixel: 001 (center)
...
+identificationInfo (basic information about
data)
        MD_DataIdentification
                …
        spatialRepresentationType: 002 (code
        for grid data)
        spatialResolution: (scale 1:1000)
                MD_Resolution
                equivalentScale:

        MD_ReprezentativeFraction
                        denominator:1000
        …
        geographicBox:
                EX_GeographicBoundingBox
                (bounding box of data in
                degrees)
                …
...
```

Listing 1. Metadata description of raster data

An example of satellite imagery description using ISO 19115 is shown in listing 1. It shows the spatial representation information which includes grid spatial

representation and can be divided in georectified and georeferencable grid. Georectified grid is a grid whose cells are regularly spaced in a geographic or map coordinate system defined in the spatial referencing system so that any cell in the grid can be geolocated given its grid coordinate and the grid origin, cell spacing, and orientation. Georeferencable grid is a grid with cells irregularly spaced in any given geographic or map projection coordinate system, whose individual cells can be geolocated using geolocation information supplied with the data but cannot be geolocated from the grid properties alone. Spatial representation information includes properties of the grid such as number of dimensions, axis properties, cell geometry, availability of check points and transformation parameters, etc. The following listing shows extract of metadata for gridded data in ISO 19115, including information about reference system, bounding box, distribution and spatial representation information which gives details of the grid.

This sort of information enables discovery and retrieval of data according to title, abstract, keywords, spatial and temporal extent, categories, themes, etc. It answers the ”what, where, when, why, who, and how” questions about geospatial resources. ISO 19115 also specifies Core metadata set which is a basic minimum number of metadata elements that should be maintained for a dataset in order to identify a dataset for catalogue purposes. It includes mandatory metadata elements as well as recommended optional elements which will increase interoperability, allowing users to understand the geographic data and the related metadata provided by either the producer or the distributor. Metadata is stored and accessed through metadata catalogues [10] that are usually implemented according to OGC Catalogue Services (CAT) specification [11] by OpenGIS consortium. This specification can be implemented using different information models among which is ISO 19115 (OpenGIS Catalogue Services ISO Metadata Application Profile [12]).

Spatial metadata can also be expressed using OASIS ebXML Registry Information Model [13]. In order to support this information model OGC has developed ebRIM profile of CAT [14] which uses ebRIM as catalogue information model over standard OGC CAT interface (ebRIM profile for Earth Observation Products is specified in [15]). This information model specifies how catalogue content is structured and interrelated. It constitutes a public schema for discovery and publication purposes. The ebXML Registry is capable of storing any type of electronic content such as XML documents, text documents, images, sounds and videos. The ebRIM uses several standard classification schemes as a mechanism to provide extensible enumeration types which are used to create classifications or ontologies for the catalogue content. The ebRIM information model is a general and flexible one with several extensibility points. A set of extensions that address the needs of a particular application domain or community of practice may be defined. The ebRIM is more generic and flexible than ISO 19115 and may contain various contents which is not specifically indented for geospatial data, and in that way the relationship between GIS and non-GIS systems is provided. For this reason, OGC has proclaimed it for the recommended application profile. An example of metadata in ebRIM format, namely ebRIM profile for Earth Observation (EO) Products is shown in Listing 2.

```
<adsHeader><!--Information that applies
to the entire data set-->
    <missionId>S1A</missionId> <!--
Information about sensor. Sentinel 1-->
    <productType>GRD</productType><!--
Ground Range Detected product type-->
    <polarisation>VH</polarisation>
        <!—Radar  polarisation-->
    <mode>IW</mode>    Interferometric <!-
-Wide Swath mode-->
    <swath>IW</swath>
    <startTime>2014-10-
05T16:33:15.005782</startTime><!--
Beginning of acquisition-->
    <stopTime>2014-10-
05T16:33:40.004681</stopTime><!--End of
acquisition-->
<absoluteOrbitNumber>2697</absoluteOrbitN
umber><!--Absolute orbit number considers
the orbits elapsed since the first
ascending node crossing after launch. -->
<missionDataTakeId>12340</missionDataTake
Id><!-- Mission ID-->
    <imageNumber>002</imageNumber><!--
Delivered image number. 001 is for other
image with VV polarisation (VV-VH,dual
polarisation). -->
</adsHeader>
```

Listing 2. Metadata describing Sentinel data according to ebRIM profile for EO

The difference between ISO 19115 and ebRIM metadata can be summarized as the choice between generality and simplicity. ISO 19115 defines a detailed structure of the content of metadata for geospatial data and services. This allows better interoperability, but limits metadata to what can fit inside the ISO model. On the other hand, ebRIM allows the description and storage of the various content and data structures. ebRIM is more powerful in terms of flexibility, but is less strong as the data model specification, given that the use of ebRIM model is not sufficient to agree on the representation of metadata and achieve interoperability. Therefore, the choice between ISO and ebRIM model is a choice between simplicity and consistency on the one hand, and the expressiveness and flexibility on the other. This generality and flexibility is the reason why ebRIM is recommended by OGC for storing metadata.

Metadata catalogues may facilitate retrieval of the open data. However, the retrieval of the data is only based on keyword-based search, and the part concerning the semantics of the data is still missing and the user is not able to see the details about underlying data model. Retrieval of the data should consider feature attributes which can be spatial, thematic, qualitative and temporal. Although application schema may be referenced in metadata set, the problems of heterogeneity of formats for its representation, as well the meaning of schema elements persist and therefore it is not suitable for the any kind of automatic processing.

Record orientation of catalogues as in ISO 19115, is a clear user / client paradigm but it is hard to maintain and limited for complex metadata relationships. A registry model makes catalogs easier and more flexible to maintain, but it is rather complex when exposed to the clients. ebRIM allows the classification of data and services into categories which only partially solves the problem of semantics by introducing taxonomy, but non-taxonomic relationships are hard to maintain. Possible solution for the problem is the introduction of formal ontologies, namely OWL, a semantic markup language

for the web [16]. OWL is also proposed by OGC as the information model for CAT in OWL Application Profile of CAT [17]. An example OWL class `SentinelPlatform` representing Sentinel platform, a subclass of SatelitePlatform is shown in Listing 3. It contains many properties including, stripmap, interferometric wide swath, wave mode, extra wide swath, etc.

```
eo:SentinelPlatform
  a owl:Class ;
  rdfs:subClassOf eo: SatellitePlatform;
  rdfs:subClassOf
    [ a owl:Restriction ;
      owl:allValuesFrom eo:WaveMode;
      owl:onProperty eo:hasWaveMode
    ] ;
  rdfs:subClassOf
   [ a owl:Restriction ;
     owl:onProperty eo:hasWaveMode;
     owl:someValuesFrom eo:WaveMode
   ]
rdfs:subClassOf
    [ a owl:Restriction;
      owl:allValuesFrom eo: Stripmap;
      owl:onProperty eo:hasStripmap
    ] ;
  rdfs:subClassOf
   [ a owl:Restriction;
     owl:someValuesFrom eo: hasStripmap;
     owl:onProperty eo: hasStripmap
   ] ...
```

Listing 3. OWL class SentinelPlatform

V. CONCLUSION

New technologies provide construction of smaller platforms and therefore the price of construction is decreasing. Furthermore, competition in space is also one of the parameters that can lower the price of commercial images. RapidEye constellation of five platforms is one example of light and cheaper platform that was sent to space. It is also possible in the future that the constellation will have much more satellites, more than 5. One platform can carry more sensors with different spatial, radiometric, spectral and temporal resolution. Example is Terra platform (NASA) which carries 5 sensors with different properties. Future work of scientist from this field is 'smaller and cheaper' and they strive towards developing such platforms (Skybox, Planet Labs) [18]. It is planned that data from these platforms be free for academic and non-profit researches.

One of the conclusion is that the EC decided to send significant number of remote sensing satellites in space. The data which will be delivered free of charge will be different (products, multispectral images, radar images) and will often be combined with other platforms such are USGS platforms.

Increasing number of sensors in space means more and more data to process, analyze and use in different fields. People need to be educated for this.

In this paper we have introduced the meaning of semantic metadata for EO products. Future work will be to clearly and fully describe, display and analyze the semantic metadata for Earth Observation products.

ACKNOWLEDGMENT

Results presented in this paper are part of the research conducted within the Grant No. 37017, Ministry of Education and Science of the Republic of Serbia.

REFERENCES

[1] http://opendefinition.org/od/ visited 01.2015.

[2] Action Plan for implementation of OGP initiative in the Republic of Serbia for 2014 and 2015.

[3] C. E. Woodcock et al. Science 320, 1011 (2008).

[4] M.A. Wulder, and N.C. Coops, "Satellites: Make Earth observations open access," *Nature*, vol. 513, pp. 30-31, September 2014.

[5] CEOS, 2014. The Committee on Earth Observation Satellite's Earth Observation Handbook, <http://www.eohandbook.com> (accessed 17.02.14).

[6] A.S. Belward, and J.O. Skoien, "Who launched what, when and why; trends in global land-cover observation capacity from civilian observation satellites" *ISPRS J. Photogram. Remote Sensing*, 2014, doi:10.1016/j.isprsjprs.2014.03.009.

[7] Technical Announcement: USGS Completes Decommissioning of Landsat5. http://www.usgs.gov/newsroom/article.asp?ID=3626#.VJfvVVAM8g

[8] J. Dwyer, and T. Loveland, "The Landsat Program: Current Status and Future Plans," Climate and Land Use Change, Earth Resources Observation and Science (EROS) Center.

[9] ISO 19115:2003 - Geographic information – Metadata, http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020

[10] Govedarica, M., Bošković (Sladić), D., Petrovački, D., Ninkov, T., Ristić, A., 2010. Metadata Catalogues in Spatial Information Systems. Geodetski list, vol.64 (87) no.4, pp. 313-334.

[11] OpenGIS Catalogue Service Implementation Specification 2.0.2, 2007. http://www.opengeospatial.org/standards/cat

[12] OpenGIS Catalogue Services Specification 2.0.2 - ISO Metadata Application Profile (1.0.0), http://portal.opengeospatial.org/files/?artifact_id=21460

[13] OASIS ebXML Registry Information Model (ebRIM), http://docs.oasis-open.org/regrep/regrep-core/v4.0/regrep-core-rim-v4.0.html

[14] CSW-ebRIM Registry Service - Part 2: Basic extension package (1.0.1), http://www.opengeospatial.org/standards/cat

[15] OGC Catalogue Services Standard 2.0 Extension Package for ebRIM Application Profile: Earth Observation Products

[16] Antoniou, G., Van Harmelen, F., 2009. Web ontology language: OWL. U: Staab, S., Studer, R. (editors), Handbook on Ontologies. Springer, pp. 91–110.

[17] OGC Catalogue Services – OWL Application Profile of CSW (0.3.0), http://portal.opengeospatial.org/files/?artifact_id=32620

[18] D. Butler, " Many eyes on Earth," *Nature*, vol. 505, pp. 143-144.

ESTA-LD: enabling spatio-temporal analysis of linked statistical data

Vuk Mijović*, Valentina Janev**, Dejan Paunović**
* School of Electrical Engineering, University of Belgrade, Institute Mihailo Pupin, Belgrade, Serbia
** University of Belgrade, Institute Mihailo Pupin, Belgrade, Serbia
{Vuk.Mijovic, Valentina.Janev, Dejan.Paunovic}@pupin.rs

***Abstract*—In the recent years, Linked Data has become widely adopted and became established in the areas of data and knowledge management. Furthermore, various open government initiatives contributed to the availability of governmental data which is largely statistical in nature and often refers to different geographical regions and points in time. However, semantic technology has not influenced spatial data management yet. In this paper we discuss the possibilities for utilizing Linked Data in this domain and argue that it would facilitate integration of geospatial data with external datasets which is cumbersome in existing GIS systems. The paper addresses modelling of statistical linked data with the focus on representing spatial and time dimensions, and describes the current prototype of the Exploratory Spatio-Temporal Analysis tool for Linked Data developed by the Institute Mihailo Pupin within the GeoKnow framework.**

I. INTRODUCTION

With the wider adoption of standards for representing and querying semantic information, such as RDF(s) and SPARQL, Semantic Web technologies gained traction in the recent years and became established in the areas of data and knowledge management [1]. This process was also supported by the advances of RDF stores which have become more robust and now offer functionalities that are similar and comparable to those of traditional databases.

Geospatial data makes for a large portion of available knowledge bases and presents a highly valuable source of information for variety of applications. It can be loaded into GIS systems, however it is quite cumbersome to integrate external datasets into them and thereby leverage additional knowledge that is available. Open Geospatial Consortium enables and provides a way to share, reuse and integrate data between different GIS systems, but this data is still isolated in the GIS realm and therefore disconnected from the Web of Data. In order to tackle this issue, the EU FP7 project GeoKnow aims to make geospatial data a first-class citizen of the Web of Data and provide tools that would enable easy integration of variety of data sources and enable decision making powered by this data.

On the other hand, in the recent years global Open Government Data (OGD) initiatives, such as the Open Government Partnership[1], have helped to open up governmental data for the public, by insisting on opening non-sensitive information, such as core public data on transport, education, infrastructure, health, environment, etc. The vision for ICT-driven public sector innovation [2] refers to the use of technologies for the creation and implementation of new and improved processes, products, services and methods of delivery in the public sector. These efforts contributed to the availability of large volumes of public sector information which is mostly statistical in nature and often refers to different geographical regions and points in time.

ESTA-LD (Exploratory Spatio-Temporal Analysis of Linked Data) is a tool developed within The GeoKnow projects which aims to enable exploration and analysis of spatio-temporal linked statistical data, and demonstrate that by publishing statistics as Linked Data, it is easier to integrate external data, compare different datasets, and link to non-tabular data. This tool will also demonstrate usefulness of other tools developed within the project which will be leveraged to extract and transform data from other formats, interlink and integrate it with external datasets, and finally validate and improve its quality.

Main basis of the tool is the RDF Data Cube vocabulary, a W3C recommendation for modeling statistical data as Linked Data. The vocabulary and modeling of spatial and time dimension will be discussed in Section 2. GeoKnow project and the role of ESTA-LD will be described in Section 3, while the tool itself will be elaborated in Section 4. Finally, we will provide conclusions and give insights about the future work in Section 5.

The work described in this paper builds upon and extends previous efforts elaborated in [3].

II. MODELING SPATIO-TEMPORAL DATA

In this section we will introduce the RDF Data Cube vocabulary and how it can be used to model statistical data. Afterwards, we will go further into details and discuss how to model spatial and time dimensions in a way that will denote the role of these dimensions and enable exploration and analysis across space and time.

A. RDF Data Cube vocabulary

In January 2014, W3C recommended the *RDF Data Cube* vocabulary [4] as a standard vocabulary for modelling statistical data. The vocabulary focuses purely on the publication of multi-dimensional data on the Web. It builds upon the core of the *SDMX 2.0 Information Model* [SDMX] which is the result of the Statistical Data and Metadata Exchange (SDMX[2]) Initiative established in 2001 by seven international organizations (BIS, ECB,

[1] http://www.opengovpartnership.org/

[2] http://www.sdmx.org/

Eurostat, IMF, OECD, World Bank and the UN) with the aim to introduce and support greater efficiencies in statistical practice.

The vocabulary sees a statistical data set as a collection of observations made at some points across some logical space. The collection can be characterized by a set of dimensions that define what the observation applies to (e.g. time, country) along with metadata describing what is measured (e.g. economic activity, prices), how it is measured and how the observations are expressed (e.g. units, multipliers, status). Therefore, a statistical data set can be seen as a multi-dimensional space, or hyper-cube, indexed by those dimensions. Consequently, the vocabulary refers to statistical datasets as data cubes though this name shouldn't be taken literally since it is not meant to imply that there are exactly three dimensions (there can be more or fewer) nor that all the dimensions are somehow similar in size. Explicit definition of the cube's structure is represented by a Data Structure Definition (DSD) that enables validation, visualization, discovery, and abbreviation. DSD consists of a set of dimensions, attributes and measures, where dimensions serve to identify the observations, measures are used to describe phenomena being observed, while attributes allow to qualify and interpret the observed values. The vocabulary also allows to group subsets of observations together by creating slices through the cube in which one or more dimension values are fixed. In this case, explicit structure of a slice is given by associating it with an appropriate slice key, much like DSDs are used to describe structure of datasets.

B. *Modelling hierarchucal data*

In order to formalize the conceptualization of hierarchical dimensions we can use the Simple Knowledge Organization System (SKOS)[3]. SKOS Core is a model and an RDF vocabulary for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, 'folksonomies', other types of controlled vocabulary, and also concept schemes embedded in glossaries and terminologies. Concepts represented as skos:Concept are grouped into concept schemes (skos:ConceptScheme) that serve as code lists from which the dataset dimensions draw on their values. Semantic relation used to link a concept to a concept scheme is skos:inScheme. Herein, we will present an example of coding a geographical dimension in RDF.

```
geo:RS21 rdf:type geo:Region ;
  owl:sameAs
     <http://dbpedia.org/page/%C5%A0umadija_
and_Western_Serbia> ;
  skos:broader geo:RS ;
  skos:narrower geo:RS212, geo:RS216,
geo:RS211, geo:RS215, geo:RS213, geo:RS218 ,
geo:RS214 ,geo:RS217 ;
  skos:notation "RS21"^^xsd:string ;
  skos:prefLabel "Region of Sumadija and
Western Serbia"@en , "REGION ŠUMADIJE I
ZAPADNE SRBIJE"@sr-rs .
```

C. *Modelling spatial and time dimensions*

As part of the content oriented guidelines (COG), SDMX standard defines a set of common statistical concepts and associated code lists which are meant to be reused across different datasets. These guidelines are also available in RDF, as part of an effort by the community group[4]. Resources defined therein are not a part of the Data Cube specification, however they are used by a number of existing Data Cube publications and represent a solid foundation for modeling new dimensions. Among the provided concepts are dimensions sdmx-dimension:refPeriod and sdmx-dimension:refArea which may be used as a basis for defining time and spatial dimensions respectively. For example, dimension from the COG can be used to derive a new time dimension in the following way:

```
eg:refPeriod a rdf:Property,
qb:DimensionProperty;
  rdfs:label "reference period"@en;
  rdfs:subPropertyOf sdmx-
dimension:refPeriod ;
  rdfs:range interval:Interval;
  qb:concept sdmx-concept:refPeriod .
```

Furthermore, it is convenient to be able to easily identify which dimension is the time dimension, which component represents spatial dimension, which is a primary measure and so forth. To enable this, the vocabulary allows to denote which role a dimension plays within the structure definition. Roles are encoded as subclasses of skos:Concept and associated with dimensions through the qb:concept property. In the above example eg:refPeriod is linked to the concept it represents through the qb:concept property. This concept is of type sdmx:TimeRole, thereby denoting that this dimension represents a time dimension within the structure definition:

```
sdmx-concept:refPeriod a sdmx:TimeRole.
```

Another issue to consider is the range of these two dimensions, i.e. how to encode the dimension values. One way of representing time would be to define a code list. The problem with this approach is that it is too specialized and would limit the usability. Namely, if any tool was to be used to process this data, it would need to be aware of that particular code list in order to be able to interpret the codes. However, this problem can be overcome easily since there are two standard ways of representing the time. One solution is to use the OWL time ontology which is at the moment W3C working draft, and the other is to use xsd types such as xsd:gYear, xsd:gYearMonth, xsd:date, etc. On the other hand, the most common way to refer to geographic entities is to use resources defined in the GeoNames[5] database or link to them. Since almost every geographical dataset links to GeoNames this would enable to easily acquire any additional information that is needed. For example, if in our statistical dataset the countries were

[3] http://www.w3.org/TR/2005/WD-swbp-skos-core-spec-20051102/

[4] https://code.google.com/p/publishing-statistical-data/

[5] http://www.geonames.org/

represented with, or linked to GeoNames resources, it would be possible to acquire a polygon for any country with a simple query against the LinkedGeoData endpoint, which is like many other geographical datasets linked to GeoNames.

III. GeoKnow Generator and Linked Data Stack

GeoKnow is an EU research project that was motivated by previous work on LinkedGeoData project. Within LinkedGeoData information from OpenStreetMap was made available in RDF and interlinked with GeoNames, DBpedia, and other data sources, while GeoKnow aims to complement these efforts by making geospatial data more easily accessible on the web and improving publishing, querying, interlinking and quality assessment of geospatial information that is based on Linked Data principles.

The goal of GeoKnow is to support all stages in the Linked Data lifecycle: storage, authoring, interlinking, classification/enrichment, quality assessment, evolution/repair and searching/browsing/exploration. To achieve this goal, it includes many tools, some of which existed prior to GeoKnow and are now being improved and/or extended, while some are being developed during the course of the project (which is the case with ESTA-LD). When used together, these tools ensure better quality, and more information, thus leading to better visualizations and greater possibilities. For example, let's look at one example from the perspective of ESTA-LD. This collection of tools can be used to extract and transform data from different formats to RDF, then link it to other datasets such as GeoNames, and finally validate and repair if needed, thus leading to more data being available for analysis and ensuring high quality of this data. Furthermore, as described in the previous section, links to GeoNames can be used by the tool to acquire polygons and visualize data on a map.

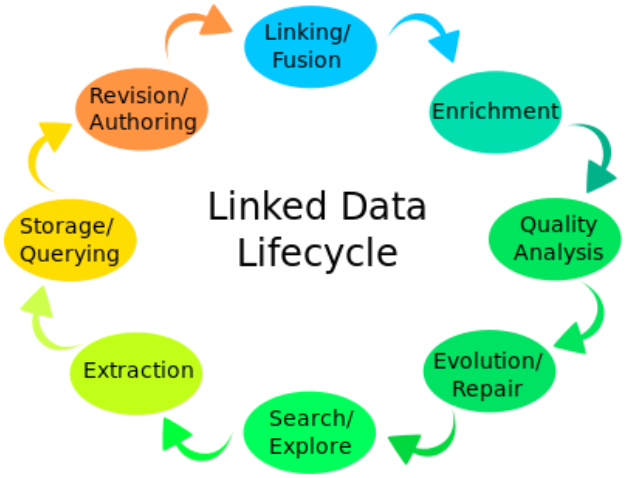


Figure. 1 Linked Data Lifecycle

Tools that are maintained and developed within GeoKnow are available as Debian packages and included in the Linked Data Stack which is a repository of Debian packages that targets Ubuntu 14.04 LTS operating system. This approach eases distribution and installation of all tools and components since the desired pieces of software can be installed with a single command without the need to deal with dependencies or configuration. The stack also includes GeoKnow Generator Workbench that integrates some of the components from the stack and provides various management functionalities such as access control and authorization, provenance, user management, or data source management, thus acting as an integrated environment that supports complete lifecycle of geospatial linked open data. We are currently in the process of creating a Debian package for ESTA-LD and making it a part of the Linked Data Stack, after which it will be integrated in the GeoKnow Generator Workbench.

IV. ESTA-LD

ESTA-LD is a tool that enables exploration and analysis of spatio-temporal linked statistical data (see Fig. 2) that is being developed within GeoKnow projects. The prototype can work on any SPARQL endpoint containing statistical data modeled with the RDF Data Cube vocabulary. It enables the user to select up to two arbitrary indicators for analysis First, it queries the endpoint for available graphs containing Data Cubes which are shown in the drop-down list on the left and when the user selects a graph, drop-down list on the right becomes populated with datasets contained in the chosen graph. Upon the selection of the dataset, its structure is analyzed and the user interface is updated accordingly. All dimensions are listed on the right side. For each dimension there is a toggle button which is used to select if the particular dimension is to be analyzed/visualized, and a drop-down list that is used to fix the dimension to a particular value.

Geographical dimension is visualized on the left side on the choropleth map. In this case, the prototype fires a query where all other dimensions are fixed to values selected in the drop-down lists on the right. This results in a set of observations for each geographical entity. Finally, the results are visualized on the map where regions for which the observed measurement is higher are depicted with a darker color then regions for which the observed measurement is lower. The map is also used to fix the geographic dimension to a particular value. Currently, the tool works with a custom defined code list of geographic areas, while in the future it will support any geographic entities linked to GeoNames.

All other dimensions are visualized on the chart positioned on the right side where up to two dimensions can be visualized/analyzed. The chart is refreshed every time a user changes the selection of dimensions to be analyzed or fixes any of the dimensions to a different value. Upon any of the mentioned changes, a query is executed in order to acquire all observations matching the selected criteria. If the user choses to visualize a single dimension, bar chart is used, while in the case of two dimensions, the component uses a histogram.

In case only a single dimension is selected for visualization and that dimension is a time dimension, a special kind of bar chart is used. This is actually a bar chart with additional controls that make it possible to select a period in time that will be shown on the chart. One of these controls is a ribbon placed below the chart. This ribbon is used to select the size and position of the time window that will be visualized. In this way it is possible to precisely set the period in time which is to be visualized, where size of the window determines duration, and its position determines the starting point. There are also pre-defined windows for periods of 1 month, 3 months, 6 months and 1 year which can be selected by clicking the dedicated buttons above the chart. At the moment, the tool is based on a custom defined code list

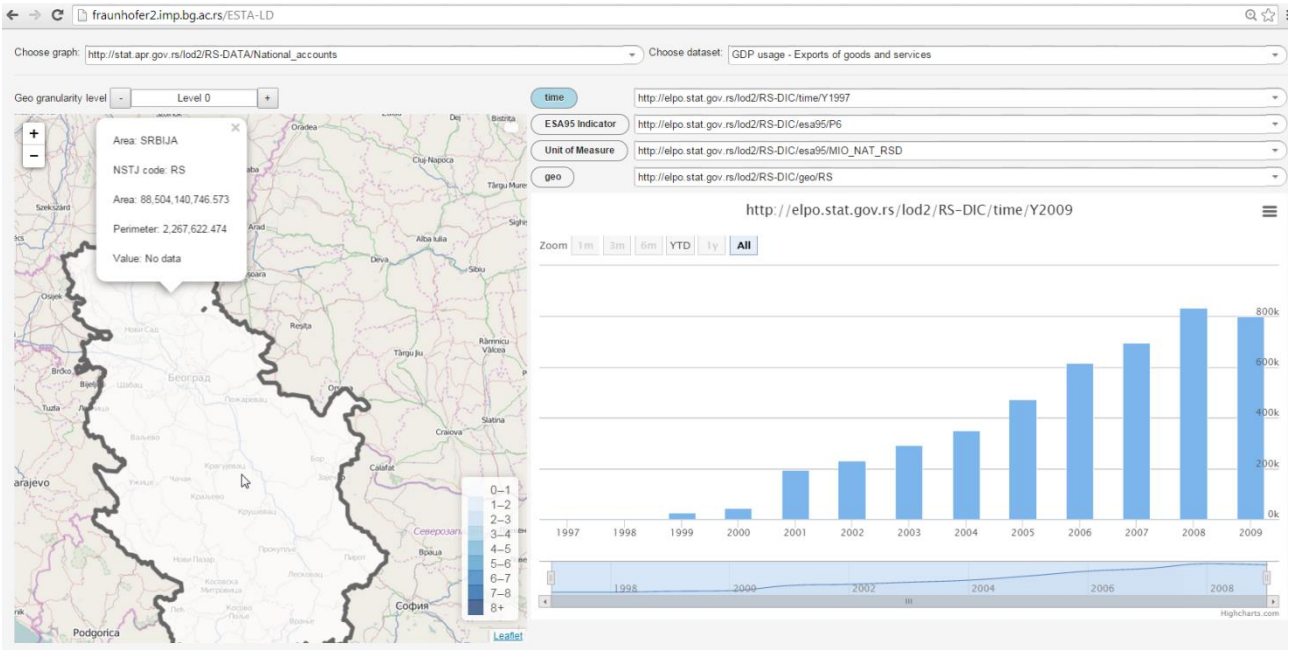


Figure 2. ESTA-LD prototype

for providing yearly and monthly data, while OWL Time and xsd types will be supported in the next version.

A. Implementation

The first prototype was developed in JavaScript and HTML5 which enabled early evaluation and testing of visualization components, while still ensuring easy integration in the *GeoKnow Generator*. Representation and interaction with geographic information were implemented using *Leaflet*[6], an open source JavaScript library for mobile-friendly interactive maps. Geographic data (such as region borders), originally available as shape files, was transformed and stored in GeoJSON format as required by Leaflet. This data is then modified using JavaScript and added to maps to create interactive visualizations. On the other hand, different statistical indicators, which are the subjects of the spatio-temporal analysis, are stored in the RDF Data Store which is queried using SPARQL query language. The actual retrieval of data from the SPARQL Endpoint was implemented using the jQuery library and its standard getJSON function. Finally, the results of the spatio-temporal analysis are visualized using *Highcharts*[7], a charting library written in pure HTML5/JavaScript, offering intuitive, interactive charts to a web site or web application.

Later on, ESTA-LD was generalized to enable the selection of indicators for analysis. In order to reuse existing Java module for querying RDF Data Cubes, new version was implemented using Vaadin[8], a Java framework for building modern web applications which also allowed for easy integration of existing functionalities. For querying the SPARQL endpoint we rely on the Sesame framework, while the user chooses the graph, dataset, and indicators to be analyzed using various Vaadin components (see Fig. 3). Since the *GeoKnow Generator* is a JavaScript web application which uses Java web servlets for the integration of Java components, and Virtuoso as an RDF store, this approach ensures straightforward integration of ESTA-LD. User interface can be easily integrated as HTML Iframe and parameters such as endpoint and initial graph to be analyzed can be specified as HTTP parameters, while the interaction and exchange of data with other components is achievable through the use of common RDF store.

B. Evaluation

Statistical data are often used as foundations for policy prediction, planning and adjustments, having a significant impact on the society (from citizens to businesses to governments) [5]. One such example is the Serbian Register of the Regional Development Measures and Incentives which is a unique, centralized electronic database of the taken measures and implemented incentives that are of significance for regional development. This register is essential for making new policies where various indicators need to be taken into

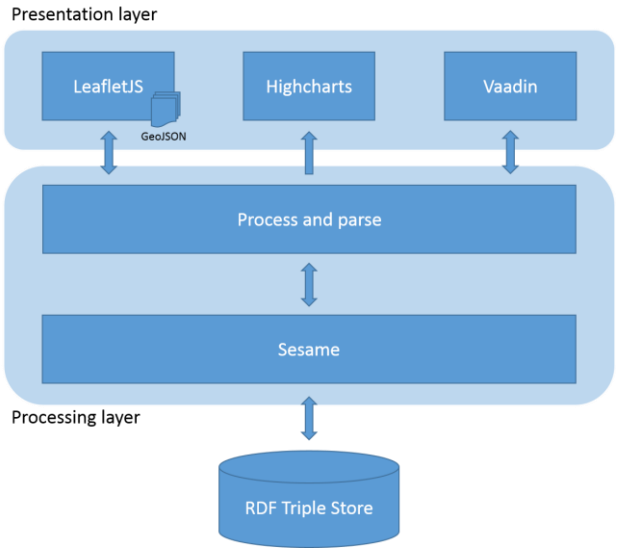


Figure 3. ESTA-LD Architecture

[6] http://leafletjs.com/
[7] http://www.highcharts.com/
[8] https://vaadin.com/home

account. In this process Linked Data technologies can be utilized to ensure interoperability and integration of data from other datasets, be it general purpose datasets such as GeoNames, or other statistical datasets such as a Register containing data from the Dissemination database of the Statistical Office of the Republic of Serbia (SORS). Data in these two registers contains both geographical and time dimensions, thus imposing challenges for analysis across geographic regions and different periods in time.

Statistics about regional development shows government investments for various purposes such as tourism, education and so forth. For each purpose, and each year, investments are captured on country, regional, and municipality level, thus allowing to evaluate the tool's ability to enable visualization and analysis across space and time where geographical data is captured on multiple levels of hierarchy. On the other hand, tourism data acquired from SORS captures different tourism indicators such as overnight stays on a monthly basis. This data allowed us to test and evaluate possibilities for analyzing different period in time. In the future, we will evaluate the benefits of analyzing data integrated from multiple sources and aggregating observations across space and time.

V. Conclusions and Future Work

This paper presented ESTA-LD, tool for exploratory spatio-temporal analysis of Linked Data. We demonstrated how statistical data can be modeled using the RDF Data Cube vocabulary and discussed different approaches to modeling spatial and time dimensions, followed by the discussion of ESTA-LD's place in the linked data lifecycle and relation to other tools for processing linked geospatial data.

Current prototype was described in detail and showed that ESTA-LD supports exploration and analysis of linked statistical data across space and time. Evaluation was primarily conducted using data published by Serbian governmental institutions that shows investments and progress of different economic indicators across geographic regions over time, thus catering to the main goal, which is to support policy makers by enabling them to utilize integration capabilities of Linked Data technologies and analyze different indicators over the integrated datasets. Currently, the tool is generic and allows exploration and analysis of arbitrary datasets and contained indicators. However, interpretation of spatial and time values is at the moment tied to custom defined code lists. Therefore, in external datasets that are not modeled with these code lists, utilizing visualization specifically tailored to spatial and time dimensions would require transformation. Consequently, the next step will be implementation of support for OWL Time and xsd types, while the generalization of the spatial dimension will be achieved by supporting geographic dimensions that contain links to GeoNames or use GeoNames resources directly. In the final stages we will take into consideration different aspects, such as scalability, flexibility and ease-of-use/friendliness.

Acknowledgment

The research presented in this paper is partly financed by the European Union (FP7 GeoKnow, Pr. No: 318159), and partly by the Ministry of Science and Technological Development of the Republic of Serbia (SOFIA project, Pr. No: TR-32010).

References

[1] J. Lehman, et al., "The GeoKnow Handbook", http://svn.aksw.org/projects/GeoKnow/Public/GeoKnow-Handbook.pdf, Accessed in December 2014.

[2] EC Digital Agenda, "Orientation paper: research and innovation at EU level under Horizon 2020 in support of ICT-driven public sector.",http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=2588, May 2013.

[3] D. Paunović, V. Janev, V. Mijović, "Exploratory Spatio-Temporal Analysis tool for Linked Data*", In Proceedings of 1st International Conference on Electrical, Electronic and Computing Engineering*, RTI1.2.1-6., June 2014, Vrnjačka Banja, Serbia.

[4] R. Cyganiak, D. Reynolds, J. Tennison, "The RDF Data Cube vocabulary", http://www.w3.org/TR/2014/REC-vocab-data-cube-20140116/, January 2014.

[5] V. Janev, V. Mijović, D. Paunović, U. Milošević, "Modeling, Fusion and Exploration of Regional Statistics and Indicators with Linked Data tools", *In Proceedings of the Third International Conference, EGOVIS 2014, Lecture Notes in Computer Science,* vol. 8650, pp 208-221, September 2014, Munich, Germany.

Exploring collaboration between public administrations through the notion of open data

Nataša Veljković*, Sanja Bogdanović-Dinić*, Leonid Stoimenov*
* University of Niš, Faculty of Electronic Engineering, Niš, Serbia
{natasa.veljkovic, sanja.bogdanovic.dinic, leonid.stoimenov}@elfak.ni.ac.rs

***Abstract*—This paper explores the possibility of extracting semantic relations from data published on government open data portals. We have introduced a relation extraction architecture that can be placed on top of the government interoperability framework, and used for creating additional relations among published datasets. If a government has not implemented interoperability framework, the architecture can be still used, since it is designed to provide the same functionality if it had open data as a data source. The proposed architecture enables extraction of semantic relations between open data and examines possible collaboration relations between data sources.**

I. INTRODUCTION

Open Government interoperability enables collaboration between governmental authorities and is often seen from three perspectives: technical, semantic and organizational, each emphasizing an important aspect of integration between data, public administrations and users. Although there are successful interoperability frameworks providing efficient mechanisms for addressing architectural issues, increasing user demands for open access to government held data are pushing governments towards embracing innovative technological approaches and stepping out of the well-known to more intelligently handle user requests. When a user places request for information on governmental open data portal, they expect to easily obtain it, without too much effort. More importantly, a user often does not precisely know what they are looking for and are keen of blaming the software for not returning the correct results. This can be solved by offering more intelligent approach to searching datasets and presenting results to the user in a way so that the search results contain not only what a user has searched for but also related datasets.

Linking datasets with each other as well as with external relevant data sources enables intelligent data search and provides more sophisticated and pertinent results. Linked open data (LOD) initiative has arisen precisely from such needs and is currently providing a powerful method for making semantic relations between datasets [1]. Being aware that the percentage of published datasets in LOD formats (RDF, XML) on open data portals is not very high, we have also explored alternative approaches, such as tags defined for each dataset and explicitly established relations, to design a method for linking information published on open data portals. In this way we have defined a new layer on top of well-established interoperability framework that utilizes implicit and explicit relationships between datasets to provide intelligent search mechanisms and responses with connected information.

In this paper we will address semantic and organizational aspects of interoperability. Semantic aspects will be reflected in analyses and implementation of possible ways for establishing relations between datasets, either via tags or explicitly defined relationships, with final goal of providing wider result context and recommendations based on user input in real-life situations. By connecting datasets, we will implicitly provide a mechanism for exploring connections between publishing agencies based on content they publish and thus tackle the organizational interoperability aspect. The main contribution of our research is a model for expanding interoperability framework with additional *Relations Extraction Layer* along with test results obtained from model application on a selected open data platform. Illustrated with realistic use cases, the presented model will confirm the importance of collaboration between publishing agencies for building semantic open data portals capable of providing clear and understandable resulting datasets.

II. INTEROPERABILITY AND OPEN DATA

The term e-government is broadly defined as the use of information and communication technologies to support business processes of government. Implementing the "e" part in the government processes brings many benefits, among which improved efficiency, transparency, accountability, improved access to public services and lower operational costs [2]. Delivering these benefits is not a simple task, especially if there is a lack of interoperability between public administration bodies. Interoperability in e-government is often expressed as the ability of information systems to support exchange of data and sharing of knowledge and information [3]. From our point of view it is possible to create an impression of interoperability, even in cases when this concept lacks implementation. To achieve such thing we will introduce the notion of open data.

Open data in the context of e-government can be defined as data published under open license on open data portals, herein available for anyone to use, reuse and freely re-distribute [4]. It can be data on enacted laws, transport data, meteorological data, statistical data or any other government held data that can be made available to the public. When publishing open data government agencies can assign an attribute to data that will represent a relationship between this data and other published data.

For example, one enacted law is related to some other laws, or poll data is related to data on citizens and data on poll places. By having this relations we can then see how the public administration agencies are related to each other. Therefore it is possible to create a new layer of interoperability in e-government that can be achieved through the notion of open data.

To explain our idea more thoroughly, we will give an overview of different layers of e-government interoperability and position a new layer that can be added independently, by using open data as the main asset of e-government openness.

A. *Interoperability layers in eGovernment*

The three most common layers of interoperability [5] in eGovernment and their goals are given in Table 1. The goal of technical interoperability is to establish a ground infrastructure for data exchange by resolving technical issues of linking diversity of computer systems and services. Interoperability at the technical layer should result with an abstraction of a communication channel, so that data can use a single level of communication [6]. The semantic interoperability layer can be established only when there is a successful information exchange, therefore, it is placed as a second layer [7]. The role of this layer is to provide exchange of meaning. By providing a common methodology, definition, structure of information and shared services for information retrieval, any person or application receiving the informaiton should be able to understands its meaning [8]. On top of interoperability layers is the organizational interoperability that aims to achieve the process agreement, or more precisely the alignment of inter- and intra-organizational processes [7].

There are other interoperability levels introduced by some authors, such as the syntactic layer [9] that is to provide the common data exchange format, or the layer of structured customer contact and support as proposed by European Public Administration Network (EPAN) [8], or the legal iteroperability layer presented by Bekkers [10]. The introduction of these layers is most concerned with what hinders interoperation. However our goal is not to discuss these interoperability layers, but to set the ground for the introduction of another perspective of the interoperability and the corresponding relations extraction layer.

TABLE I.
INTEROPERABILITY LAYERS IN EGOVERNMENT

Layer	Goal
Technical	Data exchange
Semantic	Meaning exchange
Organizational	Process Agreement

B. *The position of open data in interoperability infrastructure*

Open data portals arouse as the user interface to all publicly available data held by government agencies. As part of the global movement for achieving openness in government through opening data, procedures and processes, it became a must for a government to have an open data portal. Open data portals are structured in a way to offer data from every data provider. There are two possible scenarios that could run in the background of open data portal. In the first scenario there is successful interoperability framework adopted and implemented in the state administration. The framework provides open data from all public agencies, and these data are presented through the open data portal. Data can be searched for using various data properties, such as tags, publishing date, publishing authority, etc. Relations among different datasets may be automatically generated by the interoperability framework. Based on the data relationships to other published data user can be presented with datasets he searched for, but also with related datasets. This is an optimistic scenario that might be implemented in some more advanced e-governments. Even when there is an interoperability framework behind the e-government it is questionable, whether this framework is used as a source for open data that are to be published on the open data platforms.

The second possible scenario presumes that open data are not a product of an interoperable e-government platform, but more likely an asset of joint effort of all public administration bodies to publish their data. This presumes that each publishing organization (authority) has to enter datasets manually from the administrator's interface. With a lack of interoperability framework, publishing authority would have to manually connect datasets, it will have to specify which datasets are related and in what way are they related to the publishing dataset or existing dataset. This is more realistic scenario that is present in some less developed countries where e-government interoperability framework lacks implementation.

Both scenarios rely on the open data layer. Either as a product of interoperability framework, as in the first scenario, or as a product of the open data portal, in the second scenario, data layer represents a foundation layer for our approach to interoperability through the open data. On the top of this layer we will place the relations extraction layer that will be described in details, in the succeeding section of the paper. This layer will be responsible for creating semantic relationships between open datasets.

What we are trying to accomplish with this layer will be clarified through the answers on the following questions: How to provide related datasets to the user if there are no explicit relations between datasets? That is, how can we create a mirage of interoperability through the published open data, even when there is no underling interoperability infrastructure? How to connect datasets on the basis of their associated structure, for example by exploring their categories or tags? How to extract connections between data publishing authorities based on data relations to other data?

III. RELATIONS EXTRACTION LAYER

A. *Relations Extraction Layer Architecture*

Relations Extraction Layer (REL) represents an extension of interoperability platform. It is placed on top of the interoperable infrastructure and its role is to establish new semantic relationships between open datasets. It uses the services of underlying interoperable infrastructure, without the need of knowing the way this infrastructure operates, and applies information extraction mechanisms in order to produce new semantic-reach data collections. In such way, this layer provides connections that are essential for providing users with wider search-result context based on available data.

Fig. 1 illustrates the architecture of the proposed relations extraction layer. The foundation of relation extraction layer is open data layer that is on the other hand the top most layer of interoperable e-Government system. This layer is comprised of datasets published on government's open data portal. Datasets belong to different categories, contain different types of resources and may or may not be already inter-connected. The interoperable foundation enables defining relationships between datasets prior to their publishing online and we reference them as *explicit relationships*, but that is optional and left to publishing agencies to decide. Explicit relationships are helpful for connecting datasets, but limit information extraction possibilities to the predefined connections set only.

In order to overcome this impediment, we introduce Relationships builder layer that implements Information extraction mechanisms and applies them to underlying open datasets.

Relationships Builder examines available meta-data of published datasets and searches for possible connections based on three different features:

- Explicit relations
- Tags
- Linked data

Explicit relations origin from publishing agencies and are most likely been defined during the dataset creation process. These relations are strictly defined by referring to the source and destination datasets and relationship type (inherited from, links to, is related to, etc.). Explicit relations represent obvious feature for searching for connections between datasets, but put limitations in terms of being pre-defined.

Tags are another feature that could be utilized for establishing new connections. Tags are like explicit relations defined for each dataset during the creation process and are intended to provide mechanism for simplifying the search process. Tags could be thought of as topics on which information is published within datasets enabling user to search topics of interest. However, they are the source of implicit connections as publishing author does not know if there are or what are other datasets with the same set or a subset of tags. Tags could thus be efficiently used for grouping different and not explicitly connected datasets around the same topic or topics.

Linked open data (LOD) represent the most powerful feature for enriching data with new semantic context. LOD provides linked information from anywhere on the Web and of any type: text, structured data, image, video, and other. Publishing open data in RDF format would significantly enhance intelligent information management and enable exploitation of the Web as a platform for data and information integration [11]. CKAN opendata platform supports RDF data format for publishing datasets. Linked datasets are added using links meta-tag and they reside in the extras section of the dataset representation. Linked data are added during the dataset creation process and often reflect author's presumptions on which datasets are inter-connected.

B. *Relations Extractions Layer implementation*

The implementation of the relations extractions layer is currently in beta version. This release is depend on the

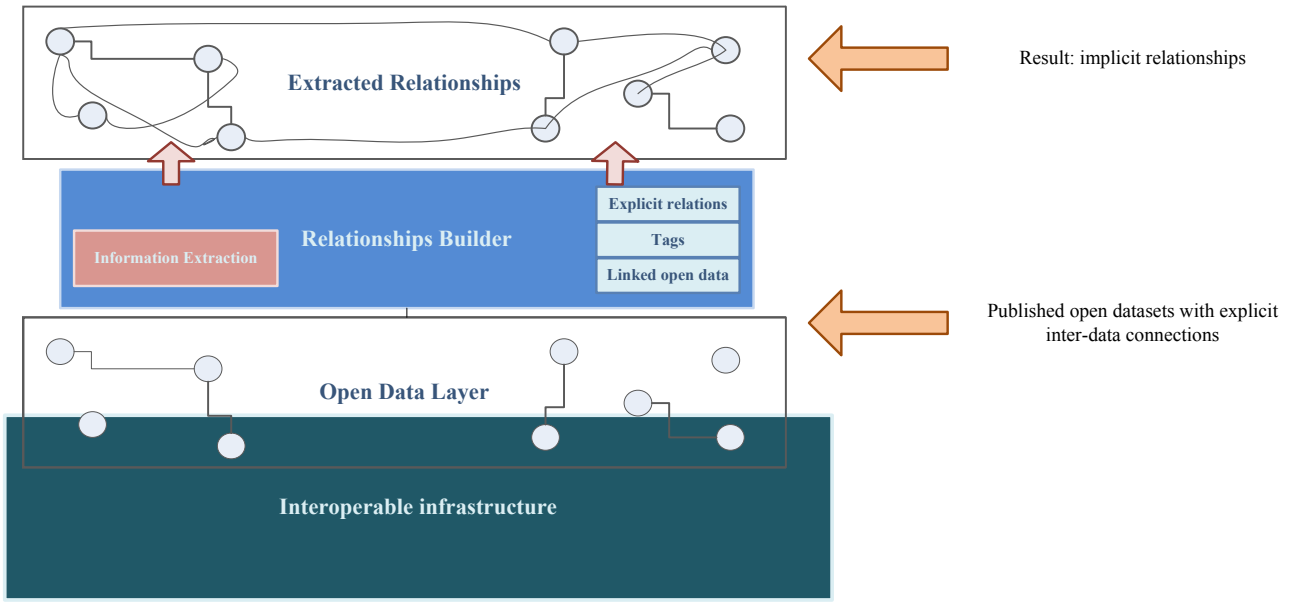


Figure 1. Relations Extraction Layer architecture

CKAN (Comprehensive Active Knowledge Network) API, but in stabile version it will be extended to support other common open data platforms. We have chosen the CKAN API since it provides the opportunity for governments to publish related datasets (packages). It also provides mechanism to automatically pull dataset dependencies when someone is downloading a dataset. Published relations can be extracted from the dataset object that we can query through API calls.

Explicit relationships provided by CKAN can be one of the following: depends_on, dependency_of, derives_from, has_derivation, child_of, parent_of, links_to, linked_from, links.

Implicit relations can be found by exploring dataset's tags. Each dataset object contains associated tags. These tags can be used for finding semantic relationship between datasets. For example we can see all the tags of a dataset and query other datasets that contain same tags. Then we can see the publishing authority of those dataset and try to find correlation between them.

REL implementation offers the user possibility to choose whether he is searching for explicit or implicit relations. And for explicit relations, we also allow user to choose the type of relation he is searching for. In the next section we will demonstrate two use-cases of relations extraction framework.

IV. USE CASE RESULTS AND DISCCUSION

On the Open Knowledge Foundation's open data portal datahub.io, there are currently 8785 datasets available. The CKAN API exposes functionality to users to retrieve all datasets names (http://datahub.io/api/1/rest/dataset). User can see all attributes of the specific dataset by requesting the dataset by its name or id (http://datahub.io/api/1/rest/dataset/abs-linked-data). CKAN will send a JSON representation of the requested dataset object (Fig. 2, Fig. 3).


```
{
   license_title: "Creative Commons CCZero",
   maintainer: "Sarven Capadisli",
   private: false,
   maintainer_email: "info@csarven.ca",
   num_tags: 21,
   id: "866ab9e5-65d1-4b6c-bf0f-c68edcbae447",
   metadata_created: "2014-08-08T09:21:03.109104",
   relationships: [ ],
   license: "Creative Commons CCZero",
   metadata_modified: "2014-08-08T10:20:35.631845",
   author: "Sarven Capadisli",
   author_email: "info@csarven.ca",
   download_url: http://abs.270a.info/sparql,
   state: "active",
   version: null,
   license_id: "cc-zero",
   type: "dataset",
 ▸ resources: [ ],
   num_resources: 5,
 ▾ tags: [
       "country-codes",
       "economics",
       "format-dcterms",
       "format-prov",
       "format-qb",
       "format-rdf",
       "format-sdmx",
       "format-skos",
       "government",
       "indicators",
       "license-metadata",
       "linked-data",
       "lod",
       "lodcloud",
       "lodcloud.candidate",
       "no-proprietary-vocab",
       "provenance-metadata",
       "published-by-third-party",
       "statistics",
       "vocab-mappings",
       "void-sparql-endpoint"
   ],
 ▾ tracking_summary: {
       total: 0,
       recent: 0
   },
```


Figure 2. JSON representation of CKAN's dataset object – tags


```
▾ organization: {
      description: "[270a Linked Dataspaces](http://270a.info/)",
      title: "270a",
      created: "2013-10-28T08:22:11.415362",
      approval_status: "approved",
      revision_timestamp: "2014-07-25T16:03:21.014562",
      is_organization: true,
      state: "active",
      image_url: http://270a.info/media/images/270a.svg,
      revision_id: "064432a8-0475-4a0d-b95f-6e61eead433d",
      type: "organization",
      id: "61fc5a3a-04c3-4fd9-a6b4-3d72ab04c606",
      name: "270a"
   },
   name: "abs-linked-data",
   isopen: true,
   notes_rendered: "<p>ABS data and metadata\n</p>",
   url: null,
   ckan_url: http://thedatahub.org/dataset/abs-linked-data,
   notes: "ABS data and metadata",
   owner_org: "61fc5a3a-04c3-4fd9-a6b4-3d72ab04c606",
   ratings_average: null,
▾ extras: {
      "links:transparency-linked-data": "99",
      triples: "2357400000",
      "links:bis-linked-data": "198",
      "links:bfs-linked-data": "99",
      "links:world-bank-linked-data": "99",
      namespace: http://abs.270a.info/dataset/,
      "links:the-eurostat-linked-data": "99",
      "links:geonames-semantic-web": "99",
      url: http://abs.270a.info/,
      "links:ecb-linked-data": "99",
      shortname: "abs.270a.info",
      "links:dbpedia": "99",
      "links:uis-linked-data": "495",
      "links:fao-linked-data": "99"
   },
   license_url: http://www.opendefinition.org/licenses/cc-zero,
   ratings_count: 0,
   title: "Australian Bureau of Statistics (ABS) Linked Data",
   revision_id: "5e9389b7-6c01-4b5d-9833-50343e1d2fec"
}
```


Figure 3. JSON representation of CKAN's dataset object - extras

For the first use-case we will try to find the correlation between publishing authorities on the basis of the dataset's tags. We will do this by extracting all tags of the current dataset and find datasets that contain any of these tags. From the framework's interface we will enter URL to portal's API and dataset name (Fig. 4).

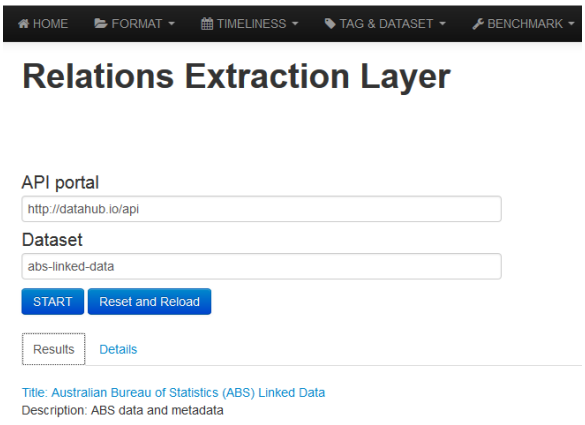


Figure 4. Interface to relation extraction framework

As a result we will receive the resource name, description and tag. By clicking on the dataset's name framework will present a tree map in the Details tab (Fig. 5) with all related datasets and their tags.

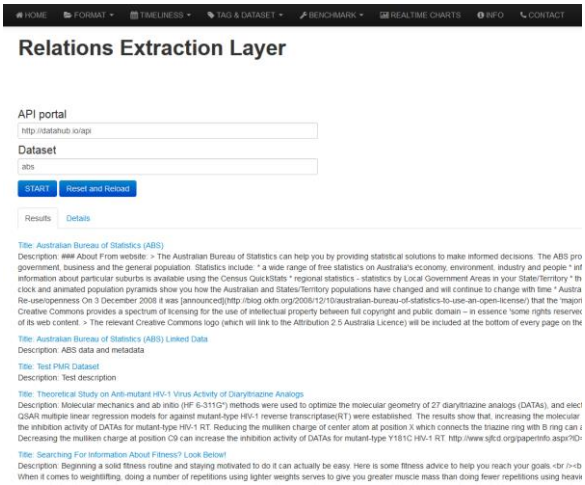


Figure 5. Relations extraction – datasets search

For the *abs-linked-data* dataset REL found 12 relevant datasets and their corresponding tags (Table II). Same tags have the same RGB color combination, so that the user can have a visual conclusion about the common tags between datasets.

From this point user can click on any listed dataset and he will receive a tree map as a visual representation of all datasets that have similar tags as the chosen dataset. In the given use-case we searched for a specific dataset by entering its' name but in the case the user does not know the dataset's name, he can enter any word and the framework will list all datasets that contain searched word in their name (Fig. 6).

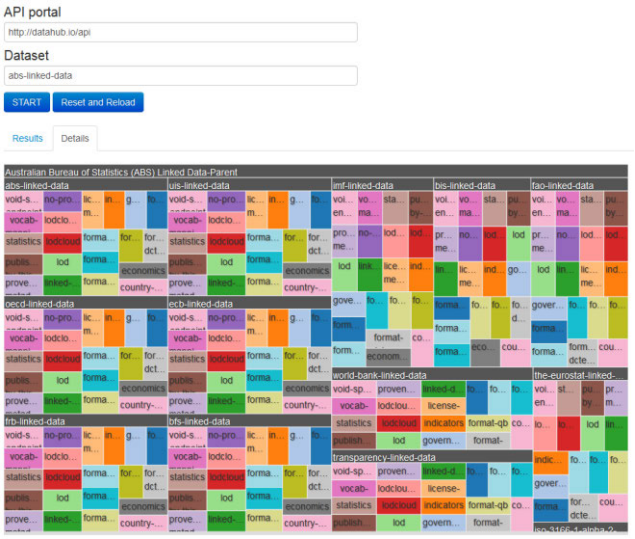


Figure 6. Relations extraction – tag level

As a second use-case we will demonstrate relations extraction layer method for extracting explicit relations. This relations are set at the time of dataset creation and they are often stored in the extras section of dataset object representation. From meta-data that are exposed for each dataset we can learn about the dataset. For extracting relations we need to look for the following meta-data: links (within extras section) and depends_on, dependency_of, derives_from, has_derivation, child_of, parent_of, links_to, linked_from (from relations section).

TABLE II.
REL IMPLICIT RELATIONS EXTRACTION - RESULTS

Dataset name	Tags
abs-linked-data-**Parent**	21 tags (country-codes, economics, format-dcterms, format-prov, format-qb, format-rdf, format-sdmx, format-skos, government, indicators, license-metadata, linked-data, lod, lodcloud, lodcloud.candidate, no-proprietary-vocab, provenance-metadata, published-by-third-party, statistics, void-sparql-endpoint, vocab-mappings)
transparency-linked-data	18 tags
imf-linked-data	21 tags (same as parent)
iso-3166-1-alpha-2-country-codes	1 tag
the-eurostat-linked-data	16 tags
ecb-linked-data	21 tags (same as parent)
frb-linked-data	21 tags (same as parent)
world-bank-linked-data	18 tags
oecd-linked-data	21 tags (same as parent)
bis-linked-data	20 tags
fao-linked-data	21 tags (same as parent)
bfs-linked-data	21 tags (same as parent)
uis-linked-data	21 tags (same as parent)

REL will search dataset object in order to see whether those predefined meta-data have value, and if so, REL will return explicitly linked datasets.

For the abs-linked-data dataset, REL extracted ten related datasets from the JSON representation of dataset object: transparency-linked-data, bis-linked-data, bfs-linked-data, world-bank-linked-data, the-eurostat-linked-data, geonames-semantic-web, ecb-linked-data, dbpedia, uis-linked-data and fao-linked-data.

Compared to the first use-case where REL used extraction by tags and found twelve datasets which are considered as implicit relations, this second method found less relations, but their significance is grater since those relations are set by the author of the dataset.

The first use-case results with seven datasets that REL also recognized using second extraction method - explicit relations extraction. Those seven common datasets are marked in Table II. There are also five more datasets found by this second method that are not listed in dataset's object returned by API call. This implies that this five datasets found by REL might be also in a strong relation with the searched dataset especially the following four datasets: imf-linked-data, oecd-linked-data, frb-linked-data and uis-linked-data, since they have 21 common tags with starting dataset. Therefore REL discovered some new relations that exist between starting dataset and found datasets. In this phase we cannot categorize discovered relation for sure and say that it is for example parent_of relation between the two datasets,

we can only say that relation exist. In the second phase REL will try to propose the relation to the user and offer the user possibility to connect the two datasets. In this use-case REL has contributed to the origin dataset since it has provided new correlated datasets.

Having relation extraction layer on top of the open data we can discuss relationships between organizations and offer possible correlations between them on the basis of relations that exist between open data.

For example if one organization publishes similar data as some other organization we can make assumption based on their published data and relations between their data that these organizations are connected in some way. Maybe one organization is operating under jurisdiction of another, or they are linked and dependent of one another in some other way. Connection on the organizational level is not yet implemented but it is one goal of the relations extraction framework. Due to the complexity in recognizing possible relations based on the small collection of parameters that can lead to conclusion about the type of relation, this problem requires more research and it will be incorporated in REL in the forthcoming research.

V. CONCLUSION

In this paper we defined an architecture that allows creating new semantic relations between data published on open data portals around the world. We have implemented beta version of the relations extraction layer and it is available on openindex.gislab.rs portal [12]. This first implementation addresses the challenge of finding related datasets and realizes two approaches described in this paper.

Through the two presented use-cases we wanted to demonstrate how REL can be used to generate meaningful new connections between datasets. The tag similarity method found five new datasets, of which four with same tags as starting dataset. For those four datasets publishing organization is 270a Linked Dataspaces that is, the same as starting dataset organization. This further implies that we can create explicit relation between starting dataset and found datasets. This relation can be a link – using link meta-tag, or even better we could propose a method that could try to identify type of relation between datasets (child_of, parent_of, depends_on, etc.). New semantic relation can be proposed on the basis of similar tags but it can also include other meta-data, such as publishing organization, data resources, data openness, etc., in order to better propose relationship type. As aforementioned CKAN offers six possible relations between datasets. For now we don't have a reliable method for recognizing what should be the type of relation between starting dataset and found datasets and so we propose link meta-tag for assigning discovered datasets and linking them to starting dataset. However our future goal is finding a proper method to propose what relationship type would be most appropriate between two datasets based on their similarity and meta-data.

VI. REFERENCES

[1] W3C. Linked Data. Retrieved from http://www.w3.org/standards/semanticweb/data

[2] N. Veljković, S. Bogdanović-Dinić, and L. Stoimenov, "Municipal Open Data Catalogues," CEDEM 2011, Edition Donau-Universität Krems, pp. 195-207

[3] Goldkuhl, G. (2008). The challenges of interoperability in e-government: Towards a conceptual refinement. Pre-ICIS 2008 SIG e-Government Workshop, Paris.

[4] N. Veljković, S. Bogdanović-Dinić and L. Stoimenov, "eGovernment Openness Index," ECEG 2011, Academic Publishing Ltd., pp. 571-577

[5] EC, The European Commission, European Interoperability Framework for Pan-European eGovernment Services, 2004.

[6] UNDP, United Nations Development Programme: e-Primers for the Information Economy, Society and Polity, "e-Government Interoperability," 2008.

[7] M. Novakouski and G.A. Lewis, "Interoperability in the e-Government Context," SEI, Carnegie Mellon University, Pittsburgh, pp. 1–35, 2012.

[8] EPAN, European Public Administration Network, "eGovernment Working Group: Key Principles of an Interoperability Architecture," Brussels, 2004, Available at http://www.epractice.eu/document/2963

[9] H. Kubicek and R. Cimander, "Three dimensions of organizational interoperability Insights from recent studies for improving interoperability frame-work," *European Journal of ePractice,* vol.6, 2009.

[10] V. Bekkers, "The Governance of Back Office Integration in E-Government: Some Dutch Experiences," in *EGOV 2005* M.A.Wimmer et al., Eds. Berlin: Springer LNCS 3591, 2005, pp. 12–25.

[11] Beckett D., McBride B. (2004, February 10). RDF/XML Syntax Specification (Revised). Retrieved from http://www.w3.org/TR/2004/REC-rdf-syntax–grammar20040210/

[12] Relation Extraction Layer implementation, http://openindex.gislab.rs/collaboration

Visual Analytics of Traffic-Related Open Data and VGI

Jan Ježek*, Karel Jedlička**, Jan Martološ***
* University of West Bohemia in Pilsen/Department of Computer Science and Engineering, Pilsen, Czech Republic
** University of West Bohemia in Pilsen/Department of Mathematics, Pilsen, Czech Republic
*** EDIP s.r.o, Pilsen, Czech Republic
jezekjan@kiv.zcu.cz, smrcek@kma.zcu.cz, martolos@edip.cz

***Abstract*—Automobile traffic problems such as car accidents and traffic congestions are touching the daily life of many people. The development of demography and the urban planning play a key role that influences the traffic situation. An easily understandable and widely accessible visualization of current and future traffic volumes can influence nowadays decisions and bring important new insights.**

Currently available open data together with VGI (volunteered geographic information) might be considered for the traffic volume prediction. In this paper, we focus on the utility of available open data and VGI (e.g. OpenStreetMap[1]) together with advanced, web-based visualization technique (based on WebGL) with the aim to offer an easy exploration and insights discovery in complex data that relates to traffic. A particular focus is given to traffic volume and movement history analysis.

I. Introduction

A good prediction of future traffic volumes and density is important information for many subjects. The traffic is influenced by various aspects such as road capacity, traffic day-time variations, locations of most visited places and events, actual and planned road constructions and weather conditions. These aspects make the gathering of relevant data a demanding and expensive process. Furthermore, an interactive visualization of an impact of a planned road construction, or future urban planning decision should be highly beneficial.

Nowadays, there are many relevant data sources available on the European level in the form of open data, that might be used for traffic analytics and predictions (see e.g. Transmodel[2], the GTFS[3] and its applications[4] or DATEXII[5]). Furthermore, there are new steadily developing web-based visualization approaches and tools that can provide advanced interactive techniques dedicated to exploratory visualization of complex problems that are related to traffic. Even though such tools and data are available for minimal costs in the public domain, their utilization for such a purpose have not yet become a common practice.

In this paper, we present a process of mining relevant parts of available open data and VGI data sources that are suitable for prediction and history analysis of traffic. In particular we focus on prediction of traffic volume (the amount of vehicles which go through a network segment in a time period) by using OpenStreetMap data together with locally and globally available demographic data such as Eurostat. We also investigate the use of historical traffic data by exploring GPS tracklogs that are provided alongside with the OpenStreetMap. The main contribution of this paper is a workflow description dedicated to harmonization and processing of relevant open data and VGI in order to derive added value information about various aspects of traffic. Furthermore we demonstrate a novel web-based interactive visualization technique for traffic data by applying the concept of multiple coordinate views. This contribution is based on intermediate results of the OpenTransportNet project.

Next chapter summarizes the related work from the field of traffic analysis as well as visualization of multivariate spatio-temporal data. Section 3 describes the overall concept of prediction of traffic volume. Section 4 describes in details the data harmonization and data processing. Section 5 outlines the visualization techniques and depicts the final user interface. Finally, the conclusion is given and the future work is mentioned.

II. Related Work

Live traffic information as well as traffic prediction are important data for many subjects. A number of different methods for (automobile) traffic forecasting have been gradually developed. These methods can be divided into:

- trend methods (analogous) which assume that the prospective volume of traffic can be derived by extrapolation of current developments - basic principles described in [1].
- synthetic methods which are based on examining patterns in the behavior of participants in the transport process and these principles also applied for a prospective period - see [2, 3].

Prognostic methods can be divided in relation to the treated area into two basic groups. These groups are:

- method of uniform growth factor, which assumes homogeneous development of the transport characteristics for the entire territory,
- mathematical model of transport network which calculate with local differences in land use - see [4].

Visualization of the history of traffic data has been explored by many different approaches. According to [5] they can be classified as: a direct visualization of raw data, a visualization of aggregated data and an extraction of derived information and its visualization. The [6] describes all parts of the problem of data processing and visual analytics of moving object data and mention

[1] http://www.opentransportnet.eu/
[2] http://www.transmodel.org/index.html
[3] https://developers.google.com/transit/gtfs/
[4] https://code.google.com/p/googletransitdatafeed/wiki/PublicFeeds
[5] http://www.datex2.eu/

various technical challenges that are faced during these tasks. One of the studies [7] describes an algorithm for an extraction of a specific points from moving object trajectories and offers a visualization technique for building a flow map. Another important aspect of discovering the semantic of the data is described in [8]. Such algorithms are usually designed to be performed upon request inside the dedicated database and provides typically aggregated values suitable for a visualization without the capabilities of live interaction.

One of the steadily developing topic of exploratory data visualization is the technique based on multiple coordinated views (MCV) given by [9]. Such a technique uses various visualization metaphors for different data types, where each visualization enables interaction (such as a filtering) that is further coordinated with another view (e.g., selecting a time interval in a bar chart triggers immediate highlighting of relevant items in a map view). For the purpose of MCV visualization several frameworks a techniques exist [10, 11, 12]. In the paper we focus on the utility of web-based visualization technique for the purpose visualization of traffic volume and traffic history.

III. TRAFFIC VOLUMES

In the scope of this paper we target on two particular tasks: analysis and visualization of predicted traffic volumes and analysis and visualization of track logs provided by moving objects (e.g. cars). These task are based on scenarios collected during the OpenTransportNet as described by [13].

As it has been mentioned earlier the traffic volume is defined as a parameter of a road network which describes the amount of vehicles which go through a network segment in a time period. We can distinguish three types of traffic volumes:

- daily traffic volume (different for each day of the week),
- annual average of daily traffic volume (AADT),
- peak hour traffic volume – in the busiest hour of the day.

In general, there are three basic types of data necessary for traffic volume calculation using a mathematical traffic mode:

- Traffic generators - demographic data about places that are usually represented as points. These points can be cities, urban districts or building blocks – it depends on the granularity of the data and the desired level of detail. These data are used for estimation of traffic flows in the network. Distinguishing between different types of places such as living, industrial, service or shopping place is useful for estimation of traffic flows direction changes in time.
- Road network - well defined and topologically correct road network is the fundamental constraining graph structure, which describes allowed movements between different places.
- Calibration measurements - physical measurements of traffic volumes (traffic censes) at particular spots of the traffic network are used for calibration of calculated volumes.

To customize and process the relevant traffic data we have collected and downloaded all relevant data sources into one common database. For such a purpose PostgreSQL with PostGIS extension database server has been set-up. The process of mining and extracting relevant data from available Open Data sources is described in upcoming sections.

A. *Traffic generators*

The volume of the generated traffic can be determined as a function of the parameter characterizing the degree of attractiveness of the area for transport. This is due to the manner and degree of land use - the population in residential areas, the number of manufacturing employees, selling space in shopping areas.

Some of these data are freely available (population - Czech Statistical Office or EUROSTAT at the European level), some data are available for individual municipalities or regions in the land use planning documentation.

B. *Route network construction*

For the model of route network various data sources might be available, however different semantics are usually used. Therefore, we build a common schema based on the semantics used in the INSPIRE directive. Afterwards we derive the mapping from the source data to the INSPIRE data model. In the section we describe the harmonization of OpenStreetMap data into that model. The harmonization principles are similar for other regional data models as well.

The OSM data model (source model) was analyzed and various data exports (namely geofabrik.de, OSM2PO and raw XML export) can be used. The OSM2PO tool, which is a PostGIS extension, was used for extracting source OSM data.

C. *Common schema and harmonization*

The INSPIRE Transport Networks data model was chosen as the harmonized data schema, because it addresses the linear topology and is compliant with the EU legislation. The INSPIRE Transportation Networks schema was analyzed and then data structures necessary for routing (RoadLink and RoadNode) were selected. Then the mapping function from OSM data to the INSPIRE Transportation Networks schema was built. For such a purpose a mapping table has been designed, where we map the OSM-based code list to Inspire-based code list. The Inspire FunctionalRoadClassValue and FormOfWayValue code list mapped to OSM are depicted in Table 1 and Table 2.

TABLE I. CONVERT TABLE FOR FUNCTIONALROADCLASSVALUE

Inspire	OSM
FunctionalRoadClass Value	OSM.roads.type
mainRoad	motorway, motorway_link, trunk, trunk_link
firstClass	primary, primary_link
secondClass	secondary, secondary_link
thirdClass	tertiary, tertiary_link
fourthClass	Residential, living_street, unclassified
fifthClass	all other values

TABLE II. COVERSION TABLE FOR FORMOFWAY .

Inspire	OSM
FormOfWayValue	OSM.roads.type
bicycleRoad	cycleway

dualCarriageway	motorway_link, trunk, trunk_link, primary_link, secondary_link, tertialy_link
enclosedTrafficArea	raceway
entranceOrExitCarPark	not a corresponding value
entranceOrExitService	not a corresponding value
freeway	not a corresponding value
pedestrianZone	not a corresponding value
motorway	motorway
roundabout	not a corresponding value
serviceRoad	not a corresponding value
slipRoad	not a corresponding value
singleCarriageway	all other values
tractor	not a corresponding value
trafficSquare	not a corresponding value
walkway	pedestrian, footway, steps, path

The data harmonisation process was done in two steps: Already mentioned import of routing data from OSM to PostGIS database was done by the OSM2PO. Afterwards the data were converted into the INSPIRE-based database schema using PL/pgSQL functions. The result is the physical schema of spatial database stored in PostGIS and depicted in Figure 1.

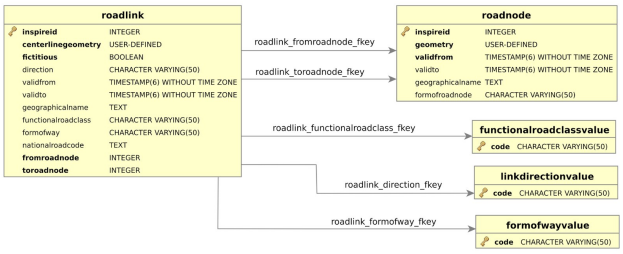


Figure 1. Physical schema of database

IV. History of Moving Objects

Another related task is to analyse and validate the predicted values by using real-world data. Ideal source for validation of traffic volumes are the data containing exact information about cars movement (GPS tracklogs - sequence of GPS coordinate measurements), however gathering such data in relevant temporal and spatial density is not an easy task. Even though such a data are collected by various corporations focused e.g. on the mobile phones industry (e.g. Google Inc.), to our best knowledge there are no Open Data with the appropriate density available for such a purpose.

A. GPS tracklogs as VGI

For the demonstration purposes, we use the GPS tracklog provided alongside with OpenStreetMap as a part of Planet.gpx initiative. Such a database offers millions of GPS tracklogs collected by volunteers from all over the word. These data consist of tracklogs provided by pedestrians, bicycles, cars as well as airplanes, ships and other types of transportation. For our demonstration purpose we extract the subset of the data that is supposed to be provided by cars. In particular we extracted the tracklogs with average speed higher than 30km/h and lower than 150 km/h and with the track length greater than 5 km. Finally we choose a time period around May 2012 and Germany as region of interest, as we found it as one of the most appropriately covered by these tracks. We extracted records consisting of these items from the database:

- Moving object position expressed as the longitude and the latitude in a geographic coordinate system.
- Timestamps of such a position.
- Id of the moving object.
- Speed of the moving object.
- Type of the road additional derived from OpenStreetMap data.

V. Data Visualization

The data are used in two use cases (traffic volumes calculation and mapping of movement history), therefore two visualization techniques which communicate the results are presented in this chapter.

A. Traffic volume visualization

Traffic volume is a spatiotemporal feature with high dynamics in time. For demonstration of its dynamics, there was created an application during the Open Data Hackaton in Jelgava (September 2014) which serves as a proof of concept. See Figure 2 for a screenshot or even the URL http://gis.zcu.cz/projekty/OTN/TrafficVolumesExample.html for the live example. The width of the RoadLink shows the amount of vehicles crossing the segment per hour. The color shows, how close is the TrafficVolume to the maximum RoadLink capacity (green = 0 % - 50 %, yellow = 50 - 70 %, red = more than 70 %. Note also the time slider, which allows the user to see the data in various times.

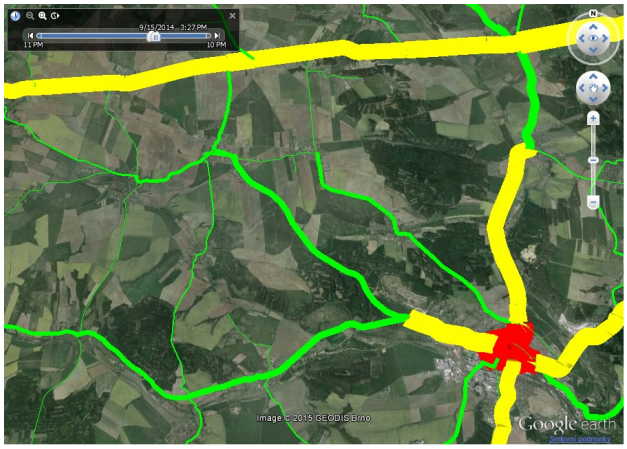


Figure 2. Traffic volume visualization

B. Visualization of GPS tracklogs

For a specified dataset we implemented a web-based multiple coordinated view visualization by using the WebGLayer and D3.js javascript libraries. In particular we configured these views:

- The number of GPS positions grouped by the rounded value of the speed displayed in the form of histogram.
- The number of GPS positions grouped by the hour of the day as a histogram
- The exact GPS position visualized as a symbol map. Symbol map visualize each position as a

small square, where the color is used to express the speed. To overcome the problem of overplotting a color transparency and blending is used.

The visualization is depicted in Figure 3. The Map is showing the positions and highlights those that are complaint with actual filters in the other views. The histograms on the right part shows the speed distribution and time of the day. The orange bars corresponds to the filter intersection, dark blue bars to the rest of the data visible in the map view, and the light blue bars to the data outside of the actual map window.

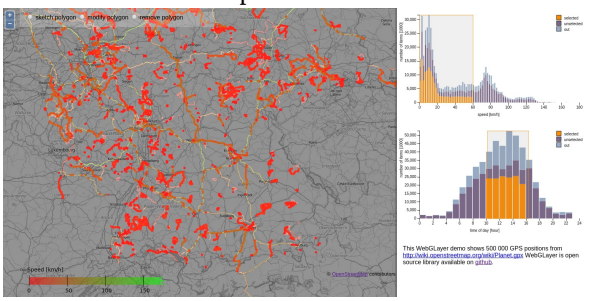


Figure 3. GPS Tracklogs visualization

The visualization is avalible as a demo of WebGLayer library at http://jezekjan.github.io/webglayer/.

VI. Conclusion

In this paper we demonstrated a combination of available open data and VGI for the purpose of car traffic analysis. For the future work we will apply the mentioned methods in the pilot cities of the OTN project. Furthermore, we target on user involvement in the OTN pilot sites in order to collect more accurate and actual VGI data. We would like to further investigate more appropriate and accurate local data sources such as live feeds and on-line information about traffic available in the form of open data. We would like also to further enrich the visualization components by additional features such as visual comparison of regions of interest.

Acknowledgment

The authors of this paper are supported by the European Union's Competitiveness and Innovation Framework Programme under grant agreement no. 620533, the OpenTransportNet project.

This action is realized by the project EXLIZ – CZ.1.07/2.3.00/30.0013, which is co-financed by the European Social Fund and the state budget of the Czech Republic.

References

[1] Benjamin, Julian. "A time-series forecast of average daily traffic volume." Transportation Research Part A: General 20.1 (1986): 51-60.

[2] Főrschner, G., "Spezifisches Verkehrsaufkommen - eine wichtige Kenngrosse fűr die Strassenverkehrsplanung," Die Strasse 10, 1/70. Berlin 1970

[3] Lerman, Steven R.,"Location, housing, automobile ownership, and mode to work: a joint choice model," Transportation Research Record 610 (1976).

[4] Florian, M., editor, "Transportation Planning Models" North Holland, Amsterdam, The Netherlands (1984).

[5] Andrienko, G., Andrienko, N., "A visual analytics approach to exploration of large amounts of movement data", Springer Berlin Heidelberg, 2008.

[6] Andrienko, G., Andrienko, N., Bak, P., Keim, D., & Wrobel, S. (2013). Visual analytics of movement. Springer Science & Business Media.

[7] Andrienko, N., & Andrienko, G. (2011), "Spatial generalization and aggregation of massive movement data", Visualization and Computer Graphics, IEEE Transactions on, 17(2), 205-219.

[8] Bogorny, V., Renso, C., Aquino, A. R., Lucca Siqueira, F., & Alvares, L. O., "CONSTAnT–a conceptual data model for semantic trajectories of moving objects." Transactions in GIS, 18(1), 66-88. Chicago, 2014.

[9] Roberts, J. C. (2007, July). State of the art: Coordinated & multiple views in exploratory visualization. In Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on (pp. 61-71). IEEE.

[10] Ježek, J., Bernard, J., Kolingerová, I.: "WebGLayer for Advanced Spatial Data Exploratory Visualization on the Web", submitted to Goinformatica, Springer, 2015.

[11] Lins, L., Klosowski, J. T., & Scheidegger, C. (2013). Nanocubes for real-time exploration of spatiotemporal datasets. Visualization and Computer Graphics, IEEE Transactions on, 19(12), 2456-2465.

[12] Liu, Z., Jiang, B., & Heer, J. (2013, June). imMens: Real-time Visual Querying of Big Data. In Computer Graphics Forum (Vol. 32, No. 3pt4, pp. 421-430). Blackwell Publishing Ltd.

[13] Kozhukh, D., Jedlička, K., Mildorf, M.,Charvát, K.,Charvát K., Jr., Martolos, J., Šťastný, J., Traffic Volumes Described on Examples in the Open Transport Net Project Pilot Regions . Submitted to Proceedings of the 18th International Conference on Information Systems for Agriculture and Forestry (ISAF) 2014, Jelgava, Latvia, ISBN 978-80-905151-2-3 (under review).

Improving geoportal information search capabilities – an approach based on semantic similarity measurement

Miloš Bogdanović*, Aleksandar Stanimirović*, Leonid Stoimenov*
* Faculty of Electronic Engineering, University of Niš, 18000 Niš, Serbia
{milos.bogdanovic, aleksandar.stanimirovic, leonid.stoimenov}@elfak.ni.ac.rs

Abstract— **In this paper we will define and describe a novel approach for improving geoportal information search capabilities. The approach we present in this paper is a part of our ongoing research regarding the development of Web-based geographic information systems. Our approach is meant to be implemented within geoportals relying on federated geographic information systems (geo-information systems, GISs) as their spatial data infrastructure. Although it can be adapted for different meta-data, our approach was intended to rely on federated GISs which utilize ontological components (ontologies) for geospatial data integration purposes. Geoportal information search improvement is performed by taking advantage of existing ontological components, in particular the sense of the ontology concept names, and matching their sense with the sense of terms extracted from a natural language description of geo-information. In this way, our approach enables searching for geo-information in a heterogeneous and distributed GIS application environment.**

I. INTRODUCTION

Since the publication of the Infrastructure for Spatial Information in the European Community (INSPIRE) directive [1], we are witnessing a constant growth in a number of implemented spatial data infrastructures (SDI) [2][3][4][5]. Accessing and sharing geospatial data assembled from heterogeneous and distributed geospatial data sources are issues that necessarily emerge during the development of spatial data infrastructures. As stated in [1], "the problems regarding the availability, quality, organisation, accessibility and sharing of spatial information are common to a large number of policy and information themes and are experienced across the various levels of public authority". Therefore, development of a spatial data infrastructure (SDI) requires means which enable sharing, access and interoperability of spatial data and services.

Spatial data access is foreseen as one of the central components in a SDI. In most cases, this component is implemented as single points of discovery and access to information within SDI – a geoportal. The common characteristic of geoportal representatives is the usage of Web GIS applications in a distributed computing environment. Contemporary geoportals rely on technologies used for the development of distributed GIS solutions. The success of a single geoportal significantly depends on the quality of a Web GIS application that represents geoportal as seen from user‟s perspective. Therefore, it is evident that there is a strong dependency between the technologies used for the development of distributed GIS and geoportals.

The ability of each user to find (discover) what he/she is searching for, in terms of displayed geo-information and maps, is considered one of the primary geoportal objectives. At a glance, this request may seem rather straightforward but is in fact highly complex and directly affects geoportal usability as seen from user‟s perspective [6][7]. Individuals who do not belong to Geographic Information System (GIS) world, also referred to as "non GIS professionals", usually expect to be capable of discovering geo-information and maps using a natural language description of the information they are interested in. The ability such as this one raises geoportal usability issues and has proven to be a difficult one to achieve using current geoportal architecture [6].

Previously reported prominent examples of geoportal development, such as ones described in [8][9][10][11], indicate that contemporary geoportals relay on the usage of metadata catalogues. Metadata catalogue is usually implemented in accordance with the Open Geospatial Consortium (OGC) standard named OpenGIS Catalogue Services Specification [12]. The implementation of OGC standards improves the overall level of structural interoperability of a GIS solution. Still, in case of a geoportal implementation scenario, there are some important requests to be implemented or functionally improved: estimate relevance percentage of geo-information discovery results, classify the result using relevance percentage or analyse relationships between words used for geospatial data discovery [13]. The described state represents the motivation for the research we are continuously conducting. A part of this research will be presented in this paper. In particular, in this paper we will an approach for improving geoportal information search capabilities. The approach we will present in this paper enables users to discover geo-information within geoportals relying on federated geographic information systems (GISs) as their spatial data infrastructure. Our approach simplifies the discovery of geo-information within geoportals by enabling users to discover geo-information simply providing a natural language of geo-information they are interested in. The discovery process is based on semantic similarity measurement between natural language of geo-information and the description of ontological components used to describe the content of federated geographic information systems (GISs).

The rest of this paper is organized as follows. Section 2 discusses related work regarding semantic similarity

measurement performed within federated geographic information systems for geo-information purposes. Also, this section discusses word sense disambiguation (WSD) methods that can be used to aid geo-information discovery process. In section 3, we describe a general architecture we envision our approach to be implemented in. Section 4 describes the approach we propose for improving geoportal information search capabilities. In section 5, we present our approach in practice by providing an example of semantic similarity determination between user-defined description and an ontology concept. Section 6 concludes with an outlook to future work.

II. RELATED WORK

In the environment of distributed geo-information sources, such as a spatial data infrastructure, users are usually provided with a single uniform access point over the refined data – a geoportal [13]. Geoportals commonly enable users discover geo-information they are interested in by implementing an information search capability. Geoportal users expect search results to be in a form of homogeneous data set(s). In order to be capable of providing users with results in such form, geoportals should rely on infrastructures capable of integrating data originating from several autonomous systems [14]. In the case we investigate in this paper, these autonomous systems are geographic information systems. Thus these infrastructures can also be observed as federated geographic information systems (federated GIS) [15].

Information integration is one of the core tasks performed by a federated GIS. To be able to perform such a task, federated GIS is in a need for a mechanism that can overcome the problem of semantic heterogeneity between different geospatial data sources [16][17]. Ontologies represent common means used to solve semantic heterogeneity problems within federated GISs [17][18]. If a federated GIS provides users with a single access point over integrated data and uses ontologies for resolving semantic heterogeneity problems, it has to solve two additional problems: perform mapping between ontologies and geospatial data sources used by individual GISs [19][20], and define discoverable Web interfaces which represent information source access points [21][22]. If a federated GIS implements means for overcoming these problems, in that case each GIS within a federated GIS can be discovered by utilizing its Web interface and mappings between ontologies and geospatial data sources can be used to retrieve integrated data. Thus, a federated GIS becomes an infrastructure which enables a geoportal to perform one of its core functionalities: provide users with ability to search through integrated data originating from heterogeneous sources.

It is our opinion that a geoportal information search capabilities can be improved in the described environment. The improvement we propose allows users to describe geo-information they are interested in using their own words, their own language. Our approach takes advantage of word sense disambiguation (WSD) algorithms as an intermediate task to aid semantic similarity measurement between the sense of the ontology concept names and the sense of terms extracted from a user-defined natural language description of geo-information.

A. Semantic similarity measurement

The notion of similarity originated in psychology. Similarity is considered one of the central theoretical constructs in psychology [23]. It can be used to perform grouping among entities and to determine if some entity categories are comparable to each other. Regarding semantic similarity measurement, it is usually performed between entity types whose representation can be highly complex. Entity type representation depends on the chosen representation language which in turn makes similarity measures difficult to compare [24].

Geographic Information Science (GIScience) has widely adopted semantic similarity measures over the past decade. One of the most widely adopted semantic similarity measures is called Matching Distance Similarity Measure (MDSM) [25]. This measure is based on Tversky's feature model [26]. It supports context theory, automatically determined weights and asymmetry. Raubal [27] suggested usage of conceptual spaces, as described in [28], to achieve cognitive semantic interoperability. Schwering and Raubal [29] proposed a method to extend current semantic similarity measures by accounting for the spatial relations between different geospatial concepts. Janowicz and Raubal argued that an affordance-based representation of the context in which similarity is measured, improves the quality of similarity measure [30]. Regarding ontologies described by description logics and used in geospatial domain, there is a number of prominent proposals suggested with aim to bring ontologies closer to existing similarity theories and semantic similarity measures [31][32][33].

B. Word Sense Disambiguation

Natural Language Processing (NLP) proclaims Word Sense Disambiguation (WSD) as one of its core tasks [34]. Word Sense Disambiguation (WSD) algorithm assigns an appropriate sense(s) to each word of a given text. WSD algorithms are divided into two classes of algorithms: supervised and unsupervised methods. A supervised WSD algorithm compares information by taking advantage of labeled training data, whereas the unsupervised method does not. According to [35], WSD methods can be classified into: path-based, information content based, gloss based and vector based methods.

In a majority of cases, external knowledge sources are considered fundamental components to perform WSD. Ontologies, glossaries, thesauri, computational lexicons, corpora of texts and other sources are often utilized as external knowledge sources. Regarding WSD methods, probably the most employed external knowledge source is WordNet [36]. “Synonym sets” (synsets) represent the main building blocks used to organize WordNet as a computational lexicon. The latest version of WordNet (3.1) is available online and it contains over 155000 terms for 117000 synsets. A “synonym set” (synset) within WordNet is a structure built up of the following components: a term (word), its class (verb, noun, adjective etc.) and connections to all semantically related words along with a brief definition („gloss“) illustrating the use of the synset members. Aside from the previously enumerated components, each synset may have semantic relations defined. A semantic relation defined for a synset can be applied to all its members. WSD methods mostly utilize the following semantic relations: hypernymy (also called kind-of or is-a), hyponymy (the inverse relations of

hypernymy), meronymy (also called part-of) and holonymy (the inverse of meronymy).

Unsupervised WSD methods can utilize semantic similarity measures used to perform word disambiguation. WordNet can be effectively used within a majority of these WSD methods. In such case, similarity between terms (synsets) is determined by utilizing semantic relations defined within WordNet. For example, path-based methods measure the length of the path between two words in a graph-like structure. In this case, WordNet can be used as a graph-like structure which provides the paths. Approaches such ones described in [37], [38] and [39] have successfully utilized WordNet as a graph-like structure to perform word similarity measurement.

III. GEOPORTAL ENVIRONMENT

The purpose of this paper is to present an approach we have developed as a part of our ongoing research regarding the development of Web-based geographic information systems. Our approach targets geoportals relying on federated geographic information systems (GISs) as their spatial data infrastructure. In particular, the approach we present in this paper relies on federated GISs which utilize ontological components (ontologies) for geospatial data integration purposes. Geoportal information search improvement is performed by taking advantage of existing ontological components. We take advantage of the sense of the ontology concept names, and match their sense with the sense of terms extracted from a natural language description of geo-information. In this way, our approach enables searching for information in a federated GIS environment, as shown in Fig. 1.

Fig. 1 represents a visualization of an environment we envision our approach to be implemented in. The main facilities of the environment are the following:

- Natural language description – users define a natural language description of geo-information they are interested in.
- Federated geo-information systems (Federated GISs) – in the context of our approach, federated GISs utilize ontologies to overcome semantic heterogeneity problems
- Ontologies – a content description of different geospatial data sources used by different GISs in a federated GIS; each federated GIS maintains its ontology within a central ontology repository.
- Ontology repository – a repository capable of storing ontologies for each GIS within a federated GIS; this component can be omitted if each part of federated GIS stores ontology locally.
- WordNet computational lexicon – a computational lexicon used as a provider of a taxonomy of terms; it is used as a knowledge source to associate the most appropriate senses with terms (words) given by the user; also, it is used as a knowledge source for semantic similarity measurement purposes.
- Search Engine – this module implements information search capability by matching federated GIS ontology concepts with user-defined geospatial data description; for this purpose, search engine utilizes WordNet computational lexicon as a knowledge source.

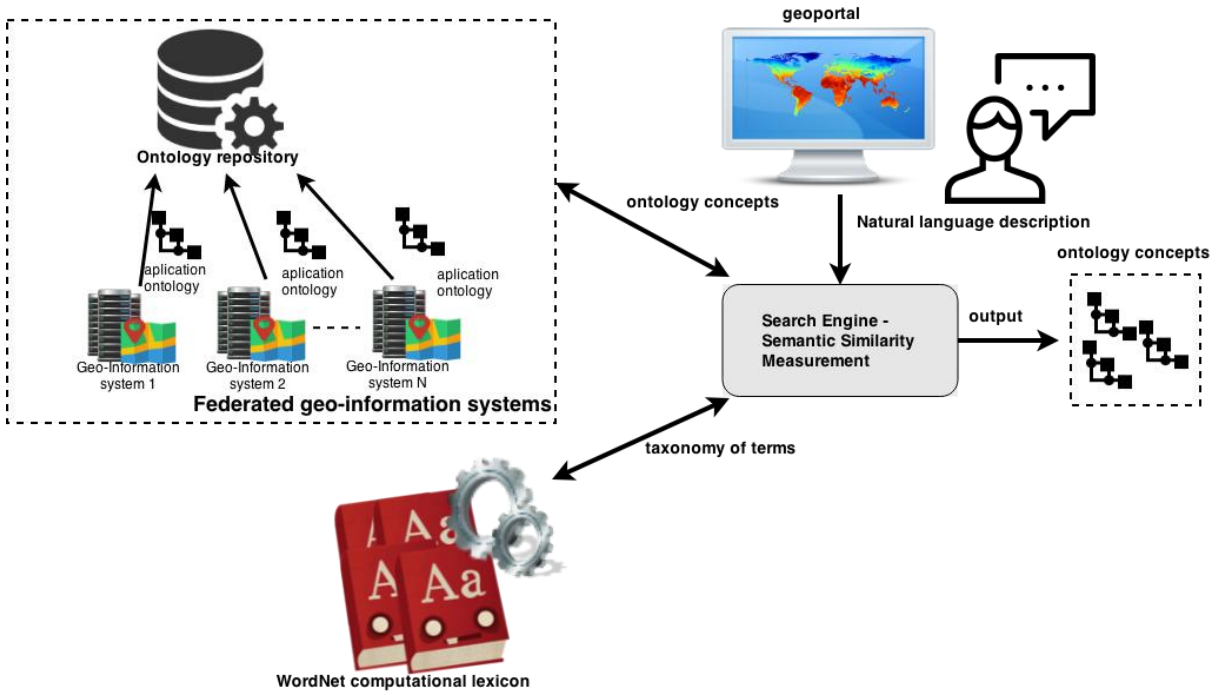


Figure 1: Geoportal in a federated GIS environment

IV. AN APPROACH FOR GEOPORTAL INFORMATION SEARCH IMPROVEMENT BASED ON SEMANTIC SIMILARITY MEASUREMENT

In this section, we will describe the approach we have developed for the purpose of improving geoportal information search capabilities. Our approach will be presented in the form of an algorithm with four sequential steps. In starts by utilizing a natural language description of geo-information given by user(s) and determines a set of concepts that belong to ontology used to describe the content of federated GIS components.

A. Disambiguate user-defined description of geospatial data

Our algorithm starts with tokenizing natural language description of geospatial information given by user(s) into a list of words. For these purposes, regular expressions are used. Afterwards, WordNet computational lexicon is utilized to identify a correct part of speech for each of the tokenized words, as shown in Fig. 2. The identification process includes the stripping of suffixes from words, as shown in Fig. 2. Suffix stripping is implemented according to algorithm defined in [40]. After the identification process is complete, words identified as nouns are extracted to a separate term set, which will be referred to as *„natural language description term set"*.

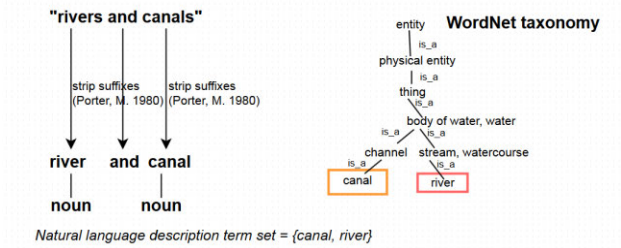


Figure 2. Disambiguation process for a user-defined description

B. Acquire and disambiguate description of ontology concepts

In most cases, ontologies concept descriptions can be acquired from *rdfs:comment* or *rdf:label* tags contained within ontology concept definitions (*rdfs:Class*). Because of performance issues, our proposal currently prefers *rdf:label* over *rdfs:comment*, since *rdf:label* contains a brief description of concept (single sentence) as opposed to *rdfs:comment* which can contain a wider description (a few sentences). The description contained within *rdfs:comment* or *rdf:label* tags is disambiguated in a way described in *Step A*. As a result of this process, a term set referred to as *„concept description term set"* is created for each concept of an ontology. As an example, Fig. 3 illustrates a disambiguation process for a description of the concept "*Word*" as defined in Suggested Upper Merged Ontology (SUMO) [41]. In this case, *rdfs:comment* was used to demonstrate the disambiguation process.

The following steps, Step C and Step D, must be repeated for each *„concept description term set"* created for each ontology concept. The same *„natural language description term set"* is used in all iterations.

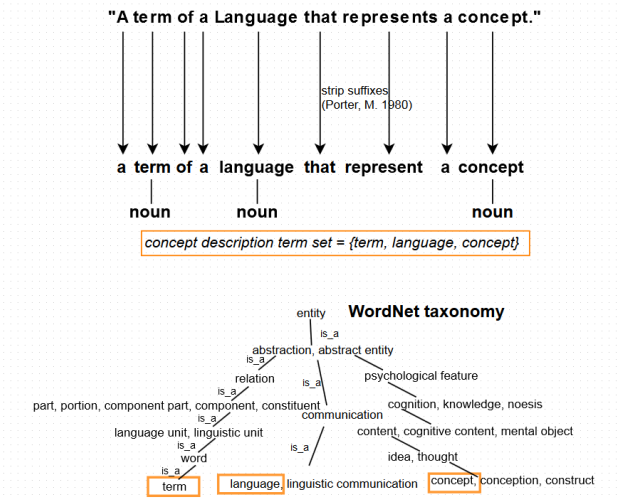


Figure 3: Disambiguated description of concept "Word" – SUMO ontology

C. Initialize and populate semantic similarity matrix

Semantic similarity matrix *S[m, n]* is created for a pair of term sets whereas each pair consists of a *„natural language description term set"* and a *„concept description term set"* created for the ontology concept whose similarity is currently being measured. *m* represents the number of terms within *„natural language description term set"* while *n* represents the number of terms within *„concept description term set"*.

In order to populate semantic similarity matrix *S[m, n]*, semantic similarity measurement is performed for each pair of terms T_{NL} and T_{CD} from *„natural language description term set"* and a *„concept description term set"*, respectively. Semantic similarity between terms T_{NL} and T_{CD} is computed according to algorithm described in [39]. Algorithm described in [39] measures the path length to the root node from the least common subsumer (LCS) of the two terms compared within a graph-based structure. In this case, WordNet computational lexicon provides paths between the observed terms. In case one of the terms or both of them do not exist in the WordNet lexicon, semantic similarity is determined according to Levenshtein distance [42]. Computed semantic similarity represents *S[i, j]*, whereas *i* represents the index of term T_{NL} within *„natural language description term set"* and j represents the index of term T_{CD} within *„concept description term set"*.

D. Calculate semantic similarity between user-defined description of geospatial data and ontology concept

The overall similarity between user-defined description of geospatial data and ontology concept is computed as an average value of similarities for each pair of terms from *„natural language description term set"* and a *„concept description term set"*, respectively (equation 1). Similarity value ranges from 0.0 to 1.0.

$$sim = \sum_{i=0}^{m-1}\sum_{j=0}^{n-1} S[i,j] \Big/ m*n \qquad (1)$$

V. AN EXAMPLE OF SEMANTIC SIMILARITY DETERMINATION BETWEEN USER-DEFINED DESCRIPTION AND AN ONTOLOGY CONCEPT

To demonstrate the algorithm described in section IV, this section will demonstrate a brief example semantic similarity computation. Let us suppose that the user is interested in finding all available geo-information regarding canals and rivers using a geoportal instance. Thus, a natural language description of geo-information given by the end user in this case could be *"rivers and canals"*. In that case, the output of the first algorithm step would be the one presented on Fig. 2. Natural language description term set consists of two terms: *"river"* and *"canal"*.

Also, let us suppose that the ontology used to describe the content of underlying geoportal infrastructure is Suggested Upper Merged Ontology (SUMO) [41]. Algorithm steps B, C and D will be demonstrated using SUMO ontology concept *WaterArea (rdfs:Class rdf:ID= "WaterArea")*. SUMO ontology describes water area as "*A body which is made up predominantly of water, e.g. rivers, lakes, oceans, etc."*. The disambiguation process regarding this description is shown on Fig. 4. As a result, concept description term set consists of five terms: *"body"*, *"water"*, *"river"*, *"lake"* and *"ocean"*.

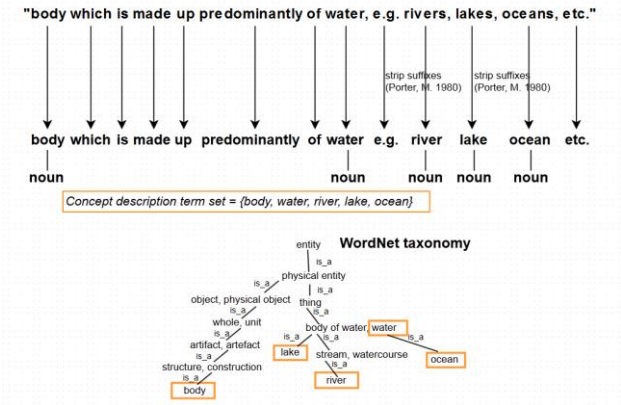


Figure 4. Disambiguation process for SUMO ontology concept *"WaterArea"*

According to algorithm Step C, semantic similarity matrix *S[2, 5]* is created and populated with semantic similarity measurement results for each pair of terms from natural language description term set and concept description term set. The resulting matrix is shown on Fig. 5.

	body	*water*	*river*	*lake*	*ocean*
river	0.5	0.8	1.0	0.73	0.73
canal	0.71	0.8	0.67	0.73	0.73

Figure 5. Similarity matrix S

The overall similarity between user-defined description of geospatial data and SUMO ontology concept named *WaterArea* is computed to 0.74 which indicates a high level of possibility that *WaterArea* ontology concept is related to geo-information the user is searching for.

VI. CONCLUSION

For the purpose of improving geoportal information search capabilities, this paper proposes an approach based on semantic similarity measurement which utilizes existing ontological components of federated GISs. The approach core is a process of computing semantic similarity between terms extracted from a natural language description of geo-information, defined by the end users, with the terms extracted from the description of ontology concepts. This process uses WordNet computational lexicon as a knowledge base for semantic similarity measurement. As a result, the approach presented in this paper outputs a set of ontology concepts that describe geospatial content within federated GISs.

Core benefits we envisioned to bring through this proposal can be summarized as follows:

- Simplified geoportal information search capabilities using natural language description of geo-information.
- The described algorithm can be implemented as a part of existing software components or as a separate Web service. Thus, it becomes independent of federated GIS‟s tier that implements geo-information access capability.
- Ontologies are used in their original form, as implemented by federated GIS developers. Therefore, our approach retains the possibility to perform reasoning over existing ontologies.

Although presented approach can be easily implemented, additional efforts should be made to make it omni-implementable. Special attention should be devoted to processing complex text expressions along with ability to utilize other types of semantic descriptions instead of ontologies.

ACKNOWLEDGMENT

Research presented in this paper was funded by the Ministry of Science of the Republic of Serbia, within the project "Technology Enhanced Learning", No. III 47003.

REFERENCES

[1] The European Parliament and The Council of the European Union, Directive 2007/2/EC of The European Parliament and of The Council, of 14 March 2007: establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), 2007

[2] Department of Environment, Community And Local Government, "Irish Spatial Data Infrastructure – Implementing the INSPIRE Directive regulations in public bodies", 2011

[3] D. Vandenbroucke and D. Biliouris, "Spatial Data Infrastructures in Slovak Republic: State of play 2010", University of Leuven, 2010

[4] Lj. Živković, "Approach to Spatial Data Infrastructure Development for Spatial Planning in Serbia", REAL CORP 2013, the 18th International Conference on Urban Planning, ISBN:978-3-9503110-5-1 Rome, Italy, May 2013

[5] D. Vandenbroucke and D. Biliouris, "Spatial Data Infrastructures in Czech Republic: State of play 2010", University of Leuven, 2010

[6] T. Aditya and M-J. Kraak, "Reengineering Geoportal: Applying HCI and Geovisualization Disciplines", In Proceedings of the 11th EC-GI & GIS Workshop, Alghero, Italy, 29 June–1 July 2005.

[7] B. Resch and B. Zimmer, "User Experience Design in Professional Map-Based Geo-Portals", ISPRS International Journal of Geo-Information, 2013, 2, 1015-1037; doi:10.3390/ijgi2041015

[8] V. Bernhard, A. Richter and M. Mittlböck, "From Geoportals to Geographic Knowledge Portals", ISPRS International Journal of Geo-Information 2.2 (2013): 256-275, 2013

[9] E. Sakkopoulos, T. Mildorf, K. Charvat, I. Berzina and K. U. Krause, "Plan4All GeoPortal: web of spatial data", In Proceedings

of the 21st international conference companion on World Wide Web (pp. 279-282). ACM, 2012

[10] F. R. Salas, E. Boldrini, D. R. Maidment, S. Nativi and B. Domenico, "Crossing the digital divide: an interoperable solution for sharing time series and coverages in Earth sciences", Natural Hazards and Earth System Sciences, 12, 3013–3029, 2012, doi:10.5194/nhess-12-3013-2012

[11] L. Bernard, I. Kanellopoulos, A. Annoni and P. Smits, "The European geoportal – one step towards the establishment of a European Spatial Data Infrastructure", Computers, Environment and Urban Systems, 29 (2005) 15–31, 2005

[12] Open Geospatial Consortium, D. Nebert, A. Whiteside and P. Vretanos Eds., OpenGIS Catalogue Services Specification, v2.0.2, OGC 07-006r1, 2007, Available from: http://portal.opengeospatial.org/files/?artifact_id=20555

[13] A. Tellez-Arenas, "Best Practice Report on Geoportals", ECP-2007-GEO-317001, OneGeology-Europe, 2009

[14] D. Calvanese and G. D. Giacomo, "Data integration: a logic-based perspective", Artificial Intelligence Magazine 26 (1) , 59–70, 2005

[15] A. Sayar, "High Performance, Federated, Service-Oriented Geographic Information Systems", PhD Thesis, Department of Computer Science, Indiana University, 2009

[16] F. Hakimpour, "Using ontologies to resolve semantic heterogeneity for integrating spatial database schemata", Ph.D. Dissertation, Zurich University, Switzerland, 191pp., 2003

[17] A. Buccella, A. Cechich, and P. Fillottrani, "Ontology-driven geographic information integration: A survey of current approaches", Computers & Geosciences, 35 (4), 2009

[18] T. Gruber, "A translation approach to portable ontology specifications", Knowledge Acquisition 5 (2), 199–220, 1993

[19] W3C RDB2RDF Incubator Group, A Survey of Current Approaches for Mapping of Relational Databases to RDF, 2009, Available from: http://www.w3.org/2005/Incubator/rdb2rdf/RDB2RDF_SurveyReport.pdf [03 December 2014].

[20] W3C RDB2RDF Working Group, RDB2RDF, 2012, Available from: http://www.w3.org/2001/sw/wiki/RDB2RDF [03 December 2014]

[21] G. Meditskos and N. Bassiliades, "A combinatory framework of Web 2.0 mashup tools, OWL-S and UDDI", Expert Systems with Applications, 38, pp. 6657–6668, 2011

[22] Y. Tian, M. Huang, "Enhance discovery and retrieval of geospatial data using SOA and Semantic Web technologies", Expert Systems with Applications, Vol. 39 (16), pp. 12522–12535, Elsevier Ltd, 2012

[23] D. Medin, R. Goldstone and D. Gentner, "Respects for similarity", Psychological Review 100(2), 254-278, 1993

[24] K. Janowicz, C. Keßler, M. Schwarz, M. Wilkes, I. Panov, M. Espeter and B. Baeumer, "Algorithm, implementation and application of the SIM-DL similarity server", In F.T. Fonseca, A. Rodriguez and S. Levashkin (Eds), Proceedings of the Second International Conference on GeoSpatial Semantics (GeoS 2007). Berlin, Springer Lecture Notes in Computer Science Vol. 4853: 128–45, 2007

[25] A. Rodriguez and M. Egenhofer, "Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure" International Journal of Geographical Information Science, 18(3) 229-256, 2004

[26] A. Tversky, "Features of similarity", Psychological Review, 84(4) 327-352, 1977

[27] M. Raubal, "Formalizing conceptual spaces", In A. Varzi, L. Vieu (Eds.): Formal Ontology in Information Systems, Proceedings of the Third International Conference (FOIS 2004). Volume 114 of Frontiers in Articial Intelligence and Applications. IOS Press, Amsterdam, 153-164, 2004

[28] P. Gaerdenfors, "Conceptual Spaces - The Geometry of Thought", Bradford Books, MIT Press, Cambridge, MA, 2000

[29] A. Schwering and M. Raubal, "Spatial relations for semantic similarity measurement", In J. Akoka, S. Liddle, I.Y. Song, M. Bertolotto, I. Comyn-Wattiau, W.J. vanden Heuvel, M. Kolp, J. Trujillo, C. Kop and H. Mayr, (Eds.): Perspectives in Conceptual Modeling: ER 2005 CoMoGIS Workshop, Klagenfurt, Austria. Volume 3770 of Lecture Notes in Computer Science. Springer, Berlin, 259-269, 2005

[30] K. Janowicz and M. Raubal, "Affordance-based similarity measurement for entity types", In: 5th Conference on Spatial Information Theory (COSIT 2007). Lecture Notes in Computer Science, Springer, 2007

[31] A. Borgida, T. Walsh and H. Hirsh, "Towards measuring similarity in description logics", In: Proceedings of the 2005 International Workshop on Description Logics (DL2005). Volume 147 of CEUR Workshop Proceedings, CEUR, Edinburgh, Scotland, UK 2005

[32] C. d'Amato, N. Fanizzi and F. Esposito, "A dissimilarity measure for ALC concept descriptions", In: Proceedings of the 2006 ACM Symposium on Applied Computing (SAC), Dijon, France, 2006

[33] K. Janowicz, "Sim-dl: Towards a semantic similarity measurement theory for the description logic *ALCNR* in geographic information retrieval", In R. Meersman, Z. Tari, P. Herrero, al., e., (Eds.), SeBGIS 2006, OTM Workshops 2006. Volume 4278 of Lecture Notes in Computer Science. Springer, Berlin (2006)

[34] R. Navigli, "Word sense disambiguation: A survey", ACM Computing Surveys (CSUR) Surveys, Volume 41 Issue 2, Article No. 10, ACM New York, NY, USA, 2009

[35] T. Zesch, and I. Gurevich, "Wisdom of Crowds Versus Wisdom of Linguists: Measuring the Semantic Relatedness of Words", Natural Language Engineering, 16(01), 1351-3249, 2010

[36] C. Fellbaum, "WordNet: An Electronic Lexical Database (Language, Speech, and Communication)", The MIT Press, 1998

[37] R. Rada, H. Mili, E. Bicknell, and M. Blettner, "Development and Application of a Metric on Semantic Nets. Systems, Man and Cybernetics", IEEE Transactions, 19(1), 17-30, 1989

[38] C. Leacock and M. Chodorow, "Combining Local Context and WordNet Similarity for Word Sense Identification", In C. Fellbaum (Ed.), WordNet: An Electronic Lexical Database (pp. 305-332). MIT Press, 1998

[39] Z. Wu and M. S. Palmer, "Verb Semantics and Lexical Selection" Proceedings of the 32th Annual Meeting on Association for Computational Linguistics, (pp. 133-138). Las Cruces, New Mexico, 1994

[40] M. Porter, "An algorithm for suffix stripping", Program (Automated Library and Information Systems), 14(3):130-137, 1980

[41] Institute of Electrical and Electronics Engineers (IEEE), Suggested Upper Merged Ontology (SUMO), Available from: http://www.adampease.org/OP/ [03 December 2014]

[42] V. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals", Soviet Physics-Doklady, Vol. 10, No. 8, 707–710. Original in Russian in Dokl. Akad. Nauk SSSR 163, 4, 845–848, 1965.

Design of Geospatial Benchmarking System and Performance Evaluation of Virtuoso and PostGIS

Mirko Spasić*
* Openlink Software, Belgrade, Serbia
mspasic@openlinksw.com

***Abstract*: A growing number of storage systems have started to support the SPARQL query language for RDF and the SPARQL Protocol for RDF. As SPARQL is more and more accepted by community, there is a growing need for benchmarks that will compare these systems. In this paper we present the setup and configuration of the GeoKnow Benchmarking laboratory and benchmarking methodology. This system is used in the FP7 project Geoknow to summarize the performance evaluation of the geospatial tools used in the project. The benchmark is applicable to DBMS dealing with relational data, and RDF, as well. Comparison results are presented running the benchmark against Virtuoso (SPARQL & SQL) and PostGIS, both hosting OSM data.**

I. INTRODUCTION

Many information systems contain some kind of geographical information. The value of this information is not the mere presence of additional columns with longitude and latitude of points or polygons. It also poses an important cue that makes data consumers relate to data, especially when the geographic data is used in data visualization, such as a map. An information system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data is called a Geographical Information System (GIS). GIS applications are tools that allow users to create interactive queries (user-created searches), analyze spatial information, edit data in maps, and present the results of all these operations.

Geographic information management is generally well-understood in data management. Many relational database systems support geographical data, sometimes by incorporating multi-dimensional indexing structures like the R-tree [1], or using simple B-trees [2]. Also, in RDF data management, many RDF stores support spatial data management, providing functions to test geospatial predicates. The Open Geospatial Consortium proposed the GeoSPARQL standard[1] to unify the specific system extensions.

In this paper we will specify the setup and configuration of the GeoKnow Benchmarking laboratory. For that purpose we could use some of the well-established benchmarks. There have been numerous benchmarks oriented towards DBMS performance in a variety of application areas. Perhaps the most famous one, TP1 is oriented toward business data processing, and has spawned a collection of derivative benchmarks, the most recent being TPC-A, TPC-B and TPC-C[2]. These benchmarks represent the typical needs of a transaction processing user of a DBMS. But, there is a broad application area, namely engineering and scientific databases, that has special needs not addressed by any of the above benchmarks. We have opted to use a geospatial benchmark built in the FP7 project LOD2[3], as a starting point, because it is more focused in addressing practical challenges in the Geo Browsing components, as developed by the University of Leipzig[4]. This benchmark emulates heavy drill-down style online access patterns and accessing large volumes of thematic data. We continued to develop and improve this benchmark. This improvement is primarily related to the expansion of the benchmark, in order to make it employable not only to RDF data, but to relational data as well. Furthermore, this will open opportunities for performance comparisons between RDF and relational spatial data management systems. This enhanced version of the benchmark will be called FacetBench. The intent is to run this benchmark against the planet-wide OpenStreetMap (OSM[5]) dataset in PostgreSQL and Virtuoso. With Virtuoso we will also compare scale-out and single server versions.

II. DATASETS AND DBMS'ES

A. Datasets

OpenStreetMap (OSM) is a collaborative project that creates and distributes free geographic data for the world. It has grown to around 300000 contributors. OSM data format has four core elements, also known as data primitives: nodes, ways, relations and tags. The experiments discussed in this paper are all on the September 2014 Open Street Map dataset. This OSM dataset cover the entire planet with 2.4bn of points. The zipped xml file from OSM had 36GB of data, while the size of zipped LinkedGeoData (LGD[6]) files (turtle format) is 177GB. The total count of the dataset is 37bn triples.

B. DBMS'es

There are several different database systems used by OSM users, but our focus will be put on PostgreSQL (with PostGIS extension) and Virtuoso. We have used

The research presented in this paper is financed by the European Union (FP7 GeoKnow project, Pr. No: 318159)

[1] http://www.opengeospatial.org/standards/geosparql

[2] http://www.tpc.org/default.asp
[3] http://lod2.eu/Welcome.html
[4] http://browser.linkedgeodata.org/
[5] https://www.openstreetmap.org/about
[6] http://linkedgeodata.org/About

Osmosis[7], a command line Java application for processing OSM data, more precisely to load planet dumps to PostgreSQL database. As we compare the different RDF stores in this benchmark, we could compare different RDBMS'es using the same data, and similar queries. Also, we can compare query running times in SPARQL and SQL this way. Our goal is to load the full planet dump to Virtuoso, as well. Instead of writing an extension of osmosis for loading osm files into Virtuoso, we will take a shortcut. We implemented ETL (Extract-Transform-Load) flows for migration of the OSM data from PostGIS to Virtuoso. In order to have fair comparison between relational data and RDF, we converted this dataset to RDF with Sparqlify[8] using the Linked Geodata mappings[9].

III. Query Workload

The selection of the queries in the workload is based on the common requests of a user dealing with the Linked Geodata browser [3], or any other map browser [4]. The workload mimics a browsing user in a query run. A query run, based on a random seed, deterministically picks 10 center points, and executes 12 steps, each step consisting of two queries: the facet count query and one instance (aggregation) queries. Thus the workload in total consists of 240 queries. The sequence of 12 steps is as follows:

1. display map at zoom level 0 at a center point
2. zoom to level 1 at the same center point
3. zoom to level 2 at the same center point
4. zoom to level 3 at the same center point
5. zoom to level 4 at the same center point
6. pan 1/8 width east at zoom level 4
7. zoom to level 5 at the same center
8. pan 1/4 height north at zoom level 5
9. zoom to level 6
10. pan 1/2 width west at zoom level 6
11. zoom to level 7
12. pan one height south at zoom level 7

The power query workload executes a query run directly after data load. It is immediately followed by the throughput workload. In the power workload, the queries in the query run are executed sequentially one after the other. In the throughput workload, multiple query runs (generated with different seeds), run concurrently on the system, simulating multiple users dealing with the data in the same time. The typical concurrency levels to test are 2, 4, 8 and 16. In this paper, we present only 16, because we want to compare the behavior of the systems being tested dealing with 1 user, and as many of them, e.g. 16.

A. Facet Count Query

The Linked Geodata Browser displays an overview with the count per facet (type) of the objects in the visible window. This is an aggregation query that counts all occurrences for each facet in the query window, be it a currently selected (active) facet or not. The query parameters here are the query center point (LATITUDE, LONGITUDE) and the window HEIGHT and WIDTH in degrees:

```
select ?f as ?facet count(?s) as ?cnt
where { ?s geo:lat ?a;
           geo:long ?o;
           a ?f.
        filter (?a >= LATITUDE-HEIGHT/2 &&
                  ?a <= LATITUDE+HEIGHT/2 &&
                  ?o >= LONGITUDE-WIDTH/2 &&
                  ?o <= LONGITUDE+WIDTH/2) }
group by ?f
order by desc(?cnt)
limit 50
```

The benchmark also allows reasonable query variants, for instance if the RDF database system being tested has specific geographic support, this can be used. For instance, Virtuoso v7 provides R-Tree based indexing, allowing to test spatial intersection within a radius. One must ensure that this RADIUS parameter fully encloses the query window so no results are missed. This variant will have the additional constraint in the filter clause:

```
bif:st_intersects( bif:st_geomfromtext(
             'POINT(LONGITUDE LATITUDE)',2000),
   ?p, RADIUS)
```

There is one more variant of this query, allowing us to test spatial intersection with a box, instead of filtering the result set by latitude and longitude constraints. It also uses benefits of utilization of R-Tree based indexing:

```
bif:st_intersects( bif:st_geomfromtext(
 'BOX(LONGITUDE-WIDTH/2 LATITUDE-HEIGHT/2,
      LONGITUDE+WIDTH/2 LATITUDE+HEIGHT/2)'),?p)
```

B. Instance Query

The Map displayed by the Linked Geodata Browser shows markers for all instances of the selected facets. To render a screen, the benchmark will always select 4 facets so there are four different FACET parameters, FACET1, FACET2, FACET3 and FACET4:

```
select ?s as ?instance ?f as ?facet
       ?a as ?lat ?o as ?lon
where {
  { ?s a ?f. filter (?f = lgdo:Village) } union
  { ?s a ?f. filter (?f = lgdo:Leisure) } union
  { ?s a ?f. filter (?f = lgdo:Tourism) } union
  { ?s a ?f. filter (?f = lgdo:Supermarket) } .
  ?s geo:lat ?a ;
     geo:long ?o .
  filter (?a >= LATITUDE-HEIGHT/2 &&
          ?a <= LATITUDE+HEIGHT/2 &&
          ?o >= LONGITUDE-WIDTH/2 &&
          ?o <= LONGITUDE+WIDTH/2)
}
```

The benchmark allows query variants that exploit the geographic capabilities of RDF database systems, as well. The information overload is a complication that can arise running this query on lower zoom levels. This is the case when facets have huge amounts of instances, which are impossible to display on the screen. So, this query is run only on higher zoom levels (5 and higher). Instead of this query, at zoom levels 0-4, the benchmark proposes instance aggregation queries.

C. Instance Aggregation Query

The problem mentioned in the previous chapter can be overcome by summarizing the instances geographically. The visible area of the map is divided into 40x20 conceptual square tiles, and the query selects only one

[7] http://wiki.openstreetmap.org/wiki/Osmosis
[8] http://aksw.org/Projects/Sparqlify.html
[9] https://github.com/GeoKnow/LinkedGeoData

instance per active facet inside one tile. It finds all the instances, but it selects only one instance in a tile (the "random" one). The benchmark provides the following query, and the query that takes advantage of geospatial support of RDF database system, as well:

```
select ?f as ?facet ?latlon ?cnt
where {
 {select ?f ?x ?y max(concat(xsd:string(?a),
                      " ",xsd:string(?o)))
                    as ?latlon count(*) as ?cnt
  where {
   {select ?f ?a ?o
           xsd:integer ( 20*
                 (?a - (LATITUDE-HEIGHT/2))/4.5
           ) as ?y
           xsd:integer ( 40*
                 (?o - (LONGITUDE-WIDTH/2))/9
           ) as ?x
    where {
     { ?s a ?f. filter(?f = lgdo:Village) }
     union
     { ?s a ?f. filter(?f = lgdo:Leisure) }
     union
     { ?s a ?f. filter(?f = lgdo:Tourism) }
     union
     { ?s a ?f. filter(?f = lgdo:Supermarket) }.
     ?s geo:lat ?a ;
        geo:long ?o .
     filter (?a >= LATITUDE-HEIGHT/2 &&
     ?a <= LATITUDE+HEIGHT/2 &&
     ?o >= LONGITUDE-WIDTH/2 &&
     ?o <= LONGITUDE+WIDTH/2)
    }
   }
  }
  group by ?f ?x ?y
  order by ?f ?x ?y
 }
}
```

D. SQL Queries

All of these queries in the previous three sections are in SPARQL language. In the following lines we present them in SQL (Virtuoso and PostGIS versions).

Facet Count Query (Virtuoso SQL):

```
select top 50 id_to_iri(type), count(*) as cnt
from nodes, node_types
where nodes.id = node_types.node_id and
      st_y(geom) >= LATITUDE-HEIGHT/2 and
      st_y(geom) <= LATITUDE+HEIGHT/2 and
      st_x(geom) >= LONGITUDE-WIDTH/2 and
      st_x(geom) <= LONGITUDE+WIDTH/2
group by type
order by cnt desc
```

Instance Query (Virtuoso SQL):

```
select id as instance, id_to_iri(t.type) as
facet,
       st_y(geom) as lat, st_x(geom) as lon
from nodes as n, node_types as t
where id=node_id and
      ( t.type = iri_to_id('lgdo:Village') or
        t.type = iri_to_id('lgdo:Leisure') or
        t.type = iri_to_id('lgdo:Tourism') or
        t.type = iri_to_id('lgdo:Supermarket') )
      and (st_y(geom) >= LATITUDE-HEIGHT/2 and
           st_y(geom) <= LATITUDE+HEIGHT/2 and
           st_x(geom) >= LONGITUDE-WIDTH/2 and
           st_x(geom) <= LONGITUDE+WIDTH/2)
```

Instance Aggregation Query (Virtuoso SQL):

```
select id_to_iri(t.type) as facet,
       CAST(20*(st_y(geom) -
          (LATITUDE-HEIGHT/2))/4.5 as int) as x,
       CAST(40*(st_x(geom) -
          (LONGITUDE-WIDTH/2))/9 as int) as y,
       max(st_y(geom) || ' ' || st_x(geom))
          as latlon, count(*) as cnt
from nodes as n, node_types as t
where n.id = t.node_id and
     ( t.type=iri_to_id('lgdo:Village') or
       t.type=iri_to_id('lgdo:Leisure') or
       t.type=iri_to_id('lgdo:Tourism') or
       t.type=iri_to_id('lgdo:Supermarket') )
     and (st_y(geom) >= LATITUDE-HEIGHT/2 and
          st_y(geom) <= LATITUDE+HEIGHT/2 and
          st_x(geom) >= LONGITUDE-WIDTH/2 and
          st_x(geom) <= LONGITUDE+WIDTH/2)
group by facet, x, y
order by facet, x, y
```

Facet Count Query (PostGIS):

```
select t.type, count(*) as cnt
from nodes as n, node_types as t
where n.id=t.node_id and
      st_y(geom) >= LATITUDE-HEIGHT/2 and
      st_y(geom) <= LATITUDE+HEIGHT/2 and
      st_x(geom) >= LONGITUDE-WIDTH/2 and
      st_x(geom) <= LONGITUDE+WIDTH/2
group by t.type
order by cnt desc
limit 50
```

Instance Query (PostGIS):

```
select id as instance, t.type as facet,
       st_y(geom) as lat, st_x(geom) as lon
from nodes as n, node_types as t
where n.id = t.node_id and
      ( t.type = 'lgdo:Village' or
        t.type = 'lgdo:Leisure' or
        t.type = 'lgdo:Tourism' or
        t.type = 'lgdo:Supermarket' )
      and st_y(geom) >= LATITUDE-HEIGHT/2 and
          st_y(geom) <= LATITUDE+HEIGHT/2 and
          st_x(geom) >= LONGITUDE-WIDTH/2 and
          st_x(geom) <= LONGITUDE+WIDTH/2
```

Instance Aggregation Query (PostGIS):

```
select t.type as facet,
       CAST(20*(st_y(geom) -
          (LATITUDE-HEIGHT/2))/4.5 as int) as x,
       CAST(40*(st_x(geom) -
          (LONGITUDE-WIDTH/2))/9 as int) as y,
       max(st_y(geom) || ' ' || st_x(geom)) as
latlon,
       count(*) as cnt
from nodes as n, node_types as t
where n.id = t.node_id and
     ( t.type='lgdo:Village' or
       t.type='lgdo:Leisure' or
       t.type='lgdo:Tourism' or
       t.type='lgdo:Supermarket')
     and (st_y(geom) >= LATITUDE-HEIGHT/2 and
          st_y(geom) <= LATITUDE+HEIGHT/2 and
          st_x(geom) >= LONGITUDE-WIDTH/2 and
          st_x(geom) <= LONGITUDE+WIDTH/2)
group by facet, x, y
order by facet, x, y
```

Each of these SQL queries has two more variants, with spatial intersection within a radius, and with a box, just like the SPARQL queries.

E. Query Parameters

Center point - LATITUDE and LONGITUDE: The queries in the benchmark are generated randomly. The part of map that is initially important for the queries is a bounding box containing a randomly chosen major city in Europe. The reason for this choice is there are plenty of instances provided by OSM, even on very high zoom levels.

Dimension of a screen – WIDTH and HEIGHT: The zoom level Z corresponds to a longitude width of $9/2^z$ degrees and a latitude height of $4.5/2^z$ degrees. So, at the lowest zoom level 0, the distance between the northernmost and southernmost points on the map is approximately 500km, and the distance from east to west is about 750km, depends on the longitude. At the highest zoom level 7, the dimension of the visible area is about 4km x 6km, corresponding to a small downtown area.

FACETS: This kind of queries only selects instances that belong to the 4 randomly chosen facets. In our example, the facets are Village, Leisure, Tourism and Supermarket.

IV. Benchmark Metrics

A. Page Per Second

The basic metric in the benchmark is PagePerSec, which is the average time to render a page of the Linked GeoData Browser, which is the sum of the facet count query and the instance (aggregation) query (depends on the zoom level); but this is reported in the inverse, hence PagePerSec. From a benchmark run, that executes each step 10 times, we derive an overall PagePerSec score at that step by averaging the 10 results (query latency in seconds). For multi-stream runs, we add the PagePerSec metric results for each stream to get a combined PagePerSec result. So, the benchmark tests how the system behaves when it handles one user at a time (power run - 1 stream row), of 16 users for the throughput run (16 streams row).

B. Page Per Second Per $1000 (PagePerSec/K$)

To take into account the cost of the hardware used in various implementations, we divide the PagePerSec metric by the dollar cost of the hardware and software used: PagePerSec/K$. If the RDF system is a commercial software product, the price for software must be the price in the US available to any customer (no discounts). The price quoted for hardware must be the publicly available end user price of the hardware at an online merchant at the date the benchmark was run.

C. Low Zoom, High Zoom and Total Score

Database systems perform quite differently at low zoom levels when compared to high zoom levels. For this reason, the benchmark reports two different sub-metrics: LowZoomScore and HighZoomScore. The first one is the geometric mean of step1 - step6, where the zoom levels are lower than 5. The HighZoomScore is derived from step6 - step12, where the zoom levels are 5, 6 or 7. The geometric mean of these two sub-metrics represents the main metrics in the Benchmark: TotalScore.

V. Benchmark Results

A. Virtuoso SPARQL results

Here, we will present Virtuoso SPARQL results, as this system is one of the most efficient storage system that expose SPARQL endpoints via the SPARQL protocol [5]. Virtuoso was run in cluster mode where one logical database is served by a collection of server processes (in our case, there was 4 of them) spread over a cluster of machines (2 machines). All the metrics are presented in Table I, and we can conclude that the average execution time of low zoom level queries is 1.66s, while the average of the high zoom levels is 0.34s, giving us the total average of 1.00s. These values are for the power run, and the following are for the throughput run with 16 users: low zoom level – 30.90s; high zoom level – 4.44s; total – 17.67s. For low zoom level queries, the average execution times for the power run are 18 times shorter than the execution times for the throughput run, while on the high zoom level the execution in 16 parallel streams is almost 13 times slower than the execution in power run. This is as expected, because the CPU utilization in power run reached the peak (the system being tested has 24 cores, so the CPU utilization was 2400%). Therefore, we could not expect shorter execution times in the throughput run.

B. Virtuoso SQL results

In this section, we present the same reporting template, but for the relational dump of Open Street Map with 2.4 billion of nodes. Here, we used the single Virtuoso instance, as well as the cluster configuration. The results for single instance are in the Table II.

In the power run, the average execution time of low zoom level queries is 6.46s. But, for the high zoom levels, the average is lower, as expected: 0.26s. The total average time in the power run is 3.36s. In the throughput run the average values for low zoom level queries, high zoom level queries, and the total average are 24.23s, 0.77s and 12.50s, respectively. The queries running in isolation executed more than 3 times faster. This is an expected result, as well, because the CPU utilization in power run was not so high.

TABLE I.
VIRTUOSO SPARQL RESULTS

Hardware	2x (dual Xeon E5-2630, 2.33GHz, 192GB RAM, 8 disks)														
Software	Virtuoso v7, Linux 2.6														
Price	$26000 @ November 23, 2013														
Metrics: PagePerSec PagePerSec/K$	step01 Z = 0	step02 Z = 1	step03 Z = 2	step04 Z = 3	step05 Z = 4	step06 Z = 4	Low Zoom Score	step07 Z = 5	step08 Z = 5	step09 Z = 6	step10 Z = 6	step11 Z = 7	step12 Z = 7	High Zoom Score	Total Score
1 stream	0.19532	0.47221	1.00594	1.18161	2.23065	2.19829	0.901762 0.035/K$	1.6787	1.89	3.4626	3.84468	6.38162	4.5065	3.2649 0.25/K$	1.71586 0.07/K$
16 streams	0.14148	0.53116	1.06908	1.41945	1.91938	2.10602	0.878953 0.035/K$	1.62373	2.46687	4.91135	4.67658	10.8606	7.404	4.41157 0.34/K$	1.96915 0.08/K$

Also, we present the results of the same benchmark, but running Virtuoso in cluster mode (4 processes, 2 machines). The reporting template is shown in the Table III.

In the power run, the average execution time of low level queries, high level queries and the total average are 0.46s, 0.03, and 0.24s, respectively, while in the throughput run these numbers are: 3.55s, 0.35, and 1.95. This is about 8 times slower.

C. PostGIS SQL results

In this section, we give the benchmark results of PostGIS in SQL as a point of reference. The dataset being tested is almost the same as the dataset in question in the previous section. The reporting template is shown in the Table IV.

In the power run, the average execution times of low zoom level queries, high zoom level queries, and total average are 218.96s, 7.91s and 113.43s, respectively. In the throughput run related numbers are only slightly higher (from 8% to 77%): 388.94s, 8.55s and 198.75s. This ratio is reasonable because the CPU utilization in power run in this case was so low.

D. Results Comparison

In this section, we summarize the results collected in the preceding ones. We present the comparison of these four systems separately on the power run, and on the throughput run.

In Figure 1 the power run comparison is presented. Virtuoso in both SQL and SPARQL outperformed PostGIS by large factor. Specifically, all the queries in the power run were executed 33 times slower in PostGIS than in Virtuoso SQL (single server). If we compare PostGIS with Virtuoso SPARQL, the factor will be even greater: 131 for low zoom level queries, 23 for high zoom level queries, and 113 in total. If we correlate Virtuoso SPARQL and SQL (single server), we will conclude that the relational version is slower almost 4 times on low zoom level queries, while it is faster 23% on high zoom level. In total, SQL version is slower more than 3 times. But, if we compare Virtuoso SPARQL and SQL, but with cluster configuration, we will conclude that SQL is faster more than 3 times on low zoom level, more than 13 times on high zoom level, and more than 4 times in total. Therefore, the largest factor is between PostGIS and Virtuoso SQL with cluster setting (more than 466).

TABLE II.
VIRTUOSO SQL RESULTS (SINGLE SERVER)

Hardware	2x (dual Xeon E5-2630, 2.33GHz, 192GB RAM, 8 disks)														
Software	Virtuoso v7, Linux 2.6														
Price	$13000 @ November 23, 2013														
Metrics: PagePerSec PagePerSec/K$	step01 Z = 0	step02 Z = 1	step03 Z = 2	step04 Z = 3	step05 Z = 4	step06 Z = 4	Low Zoom Score	step07 Z = 5	step08 Z = 5	step09 Z = 6	step10 Z = 6	step11 Z = 7	step12 Z = 7	High Zoom Score	Total Score
1 stream	0.0529207	0.0921005	0.203413	0.505894	0.927902	0.956663	0.276474 0.02/K$	2.02634	2.03004	5.71102	4.36872	13.7174	8.81834	4.80896 0.37/K$	1.15306 0.09/K$
16 streams	0.243775	0.360965	0.78531	2.01543	4.45029	4.52088	1.18727 0.09/K$	10.4645	10.7674	32.2244	25.8124	87.8918	54.6322	27.6458 2.13/K$	5.72914 0.44/K$

TABLE III.
VIRTUOSO SQL RESULTS (CLUSTER)

Hardware	2x (dual Xeon E5-2630, 2.33GHz, 192GB RAM, 8 disks)														
Software	Virtuoso v7, Linux 2.6														
Price	$26000 @ November 23, 2013														
Metrics: PagePerSec PagePerSec/K$	step01 Z = 0	step02 Z = 1	step03 Z = 2	step04 Z = 3	step05 Z = 4	step06 Z = 4	Low Zoom Score	step07 Z = 5	step08 Z = 5	step09 Z = 6	step10 Z = 6	step11 Z = 7	step12 Z = 7	High Zoom Score	Total Score
1 stream	0.631473	1.42531	4.32713	8.96891	15.3139	14.245	4.43334 0.17/K$	21.1416	22.4719	56.4972	50.5051	100	78.7402	46.8512 1.80/K$	14.4121 0.56/K$
16 streams	1.92231	3.21515	4.77236	7.49611	12.4098	12.818	5.71997 0.22/K$	14.4641	25.5404	67.2199	56.166	138.739	95.5892	56.0432 2.16/K$	17.9043 0.69/K$

TABLE IV.
POSTGIS SQL RESULTS

Hardware	2x (dual Xeon E5-2630, 2.33GHz, 192GB RAM, 8 disks)														
Software	Virtuoso v7, Linux 2.6														
Price	$13000 @ November 23, 2013														
Metrics: PagePerSec PagePerSec/K$	step01 Z = 0	step02 Z = 1	step03 Z = 2	step04 Z = 3	step05 Z = 4	step06 Z = 4	Low Zoom Score	step07 Z = 5	step08 Z = 5	step09 Z = 6	step10 Z = 6	step11 Z = 7	step12 Z = 7	High Zoom Score	Total Score
1 stream	0.00109	0.00589	0.01194	0.01681	0.02321	0.24638	0.00952 0.0007/K$	0.05367	0.06179	0.17985	0.18241	1.53304	1.069	0.23738 0.0183/K$	0.04754 0.0037/K$
16 streams	0.00964	0.04613	0.1151	0.19664	0.29607	0.30318	0.09842 0.0076/K$	0.74051	0.75399	5.28043	4.3347	23.3231	15.166	4.06399 0.3126/K$	0.63242 0.0486/K$

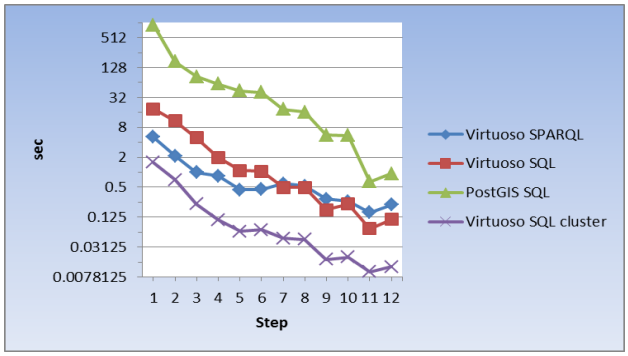


Figure 1. Power Run Comparison

In Figure 2 the throughput run comparison is shown. Virtuoso in both variants outperformed PostGIS but not with a huge factor as in the previous case. On low zoom levels, the factor was more than 16 for SQL version (single server), and 12.6 for SPARQL version; on high zoom levels, it was 11 for SQL (single), but for SPARQL it was almost 2. From Figure 2, it is obvious that PostGIS was slightly faster than Virtuoso SPARQL on the highest zoom level. Taking into account all steps from workload, PostGIS was slower almost 16 times than Virtuoso SQL (single server), and more than 11 times than Virtuoso SPARQL. Comparing Virtuoso versions, on low zoom level queries SQL version (single server) was 22% faster, while on high zoom levels it was faster almost 6 times. In total, SQL version (single server) is faster 30%. Virtuoso running on cluster was 6 times faster than running on single server, while more than 100 times faster comparing with PostGIS.

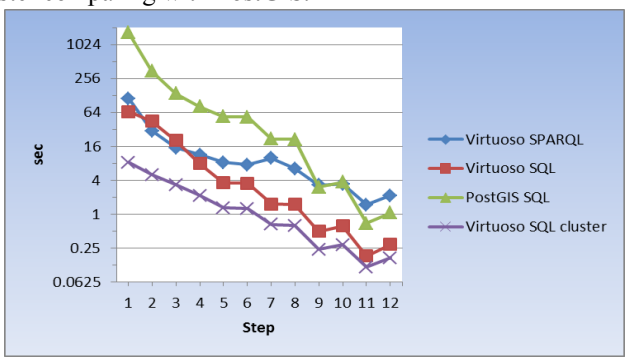


Figure 2. Throughput Run Comparison

VI. Conclusion and Future Work

We developed the first geospatial benchmark for RDF and relational data, that gives us the possibility to compare different RDF stores and RDBMSs, and compared PostgreSQL (with PostGIS extension) as one of the most common spatial database, with Openlink's Virtuoso SPARQL and SQL. This paper showed the huge victory of our database management system, as well as its scalability while comparing single server and cluster configuration.

We will integrate the tested dataset with the dataset obtained from the efforts of Linked Data Benchmark Council (LDBC[10]). In the scope of this project, Social Network Benchmark (SNB) is developed, which consists of a data generator that generates a synthetic social network. This dataset represents a snapshot of the activity of a social network for a period of time and includes entities such as Persons, Organizations, and Places. The schema also models the way persons interact, by means of the friendship relations established with other persons, and the sharing of content such as messages (both textual and images), replies to messages and likes to messages. People form groups to talk about specific topics, which are represented as tags.

We will enrich this dataset with geo-locations from LGD in the following way. For every entity that has a place assigned to it, we allocate a concrete geometry. For example, each person has its own location, e.g. home city:

```
sn:pers00000005497558391012 snvoc:isLocatedIn
<http://dbpedia.org/resource/Beijing> .
```

In this case we will assign a random LGD node representing a residential building, preferably in that city. Every university has the city where it is located, so we will enrich this with geo location, preferably real location. We will deal in the same way with locations of companies people work for. For all messages (textual and photos) persons posted, the geometries representing the exact location of a post will be assigned.

This way, we will increase the value of our benchmark, and make it more realistic. The integration provides us tremendous opportunities in terms of getting answers to variety of business questions. Some of them follow:

- We would like to know if there are differences between the places where Americans and Chinese take photos the most frequently. The possible results can show us that the first ones are interested in culture and history, while the second ones prefer entertainments, or vice versa.
- We can compare two regions (e.g. north and south of Serbia), and find out from which area people more often come to visit the other ones.
- We look for cities that are isolated, e.g. the majority of inhabitants' friends are located in radius of 10km.
- We can answer to the question, in which country people travel furthest from home to their work place.

The Virtuoso Structure-Aware RDF Store feature can optimize storage of RDF data in cases where the data exhibits regular, relational-like structure. Therefore, we will implement characteristic sets for this dataset, and gain at least 3 times acceleration in execution of SPARQL query, e.g. SPARQL will be more comparable to SQL.

[10] http://www.ldbc.eu/

References

[1] A. Guttman, "R-Trees: A Dynamic Index Structure for Spatial Searching", *SIGMOD'84, Proceedings of Annual Meeting*, pp. 47-57, 1984

[2] R. Bayer, "Binary B-trees for Virtual Memory", *Proceedings of the 1971 ACM SIGFIDET (Now SIGMOD) Workshop on Data Description, Access and Control*, pp. 219-235, 1971

[3] C. Mader, M. Martin, and C. Stadler, "Facilitating the Exploration and Visualization of Linked Data", *Linked Open Data - Creating Knowledge Out of Interlinked Data*, pp. 90-107, 2014

[4] C. Becker, C. Bizer, "DBpedia Mobile: A Location-Enabled Linked Data Browser", *Proceedings of the 1st Workshop about Linked Data on the Web (LDOW2008)*, 2008

[5] C. Bizer, and A. Schultz, "The Berlin SPARQL Benchmark", *International Journal On Semantic Web and Information Systems*, 2008

Mobile Semantic Geospatial Visualization and Exploration

Uroš Milošević*, Claus Stadler**
* Institute Mihajlo Pupin, Belgrade, Serbia
** Agile Knowledge Engineering Group, University of Leipzig, Leipzig Germany
* uros.milosevic@pupin.rs, ** cstadler@informatik.uni-leipzig.de

Abstract — **The work presented in this papers describes GEM, the first cross-platform, mobile, semantic faceted geospatial browser that fully exploits the potential of the Linked Open Data paradigm. The tool is built on top of Jassa (JAvascript Suite for Sparql Access) and offers a rich Web of Data experience by rising above the common mobile geospatial visualization limitations by relying on open, crowd-sourced and semantically linked information found in publicly available sources, such as the LOD Cloud. This information is loaded and filtered according to user's needs, on demand, in order to prevent maps from overpopulating. Finally, special attention is paid to client-side optimization to deliver both acceptable performance and comfortable user/visual experience.**

I. INTRODUCTION

The explosion of location aware technology has made the move of geographical information to their, perhaps, more natural setting, i.e. mobile devices, inevitable for any geospatial software striving to survive the demands of the ever-growing market. However, the functionality of the majority of available navigation systems is developed upon closed and proprietary solutions for both maps and software applications. Furthermore, such applications are unable to offer information specifically tailored to user's needs, and cannot be extended by third parties. Although recent attempts propose ways of overcoming some of these barriers [1, 2, 3], none leverage the full power of the Linked Data paradigm.

The design and usability choices of desktop applications make them often hard or impossible to interact with on mobile devices due to both hardware (smaller screens, lower screen resolutions, lack of buttons, less processing power etc.) and software constraints. Therefore, most existing spatial-semantic visualization and exploration applications might prove impractical for a user on the go (e.g. in a car, on a bike, on foot etc.). Also, a mobile semantic browser designed with flexibility and extensibility in mind could pave the way for semantic authoring of information on the move, which could, in turn, enable crowdsourcing of geospatial information on the Web of Data.

Our mobile spatial-semantic visualization and exploration tool aims to complement the desktop experience through a rich mobile alternative that will exploit all strengths of Linked Data and further rise above the common mobile geospatial visualization limitations by relying on open, crowdsourced and semantically linked information found in publicly available sources, such as the LOD Cloud. This information must be loaded and filtered according to user's needs, on demand, in order to prevent maps from overpopulating. Moreover, in order to reach a larger target population and spark community engagement and contribution, such a browser needs to be able to run on at least one major mobile platform (Android, iOS, Windows Phone).

In Section 2, we will first look at Jassa, the foundation library we build on top of, and the basics of semantic faceted classification and navigation. Next, in Section 3, we explore the existing and tailor-made technologies that are needed to take the semantic faceted browsing over RDF datasets to handheld devices. Section 4 showcases our take at a mobile geospatial Linked Data exploration experience, and, finally, Section 5 discusses future work and concludes this paper.

II. FOUNDATION

To provide the means to easily explore and visualize spatial content retrieved from the Web of Data, new information retrieval paradigms need to be explored. Apart from this, new techniques need to be developed to combine semantic information with geospatial data and display them with state-of-the-art map rendering tools. Moreover, to support the reuse of certain application units and separate different concerns, like browsing and presentation, a software implementing these paradigms and techniques should consist of distinguishable components.

With RDF, a data model was introduced, which enables global identification and integration of resources as well as cross-dataset interlinking. On top of RDF, SPARQL became a standard language for accessing Web databases. Yet, the development of Web applications for the exploration and visualization of SPARQL-accessible data still poses several challenges related to performance and design. The Open Source *JAvaScript Suite for Sparql Access (Jassa)* [4] is motivated by the goal of creating the generic widgets for faceted browsing of RDF data. The library is designed to tackle many of the challenges encountered during the development process.

A. Jassa

Jassa is an umbrella term for a set of three related projects. These projects are summarized as follows:

The research presented in this paper is partly financed by the European Union (FP7 GeoKnow, Pr.No: 318159), and partly by the Ministry of Science and Technological Development of the Republic of Serbia (SOFIA project, Pr. No: TR-32010).

- *Jassa Core*[2] is a project that provides a layered set of modules for: the representation of RDF and SPARQL, the execution of queries, SPARQL to JSON mapping, and most prominently, faceted search.
- *Jassa UI*[3] is a project for user interface components based on Jassa Core and the *AngularJS*[4] framework.
- *Complementary Services for Jassa*[5] is a Java project that offers server side APIs that enhance Jassa, such as a SPARQL cache proxy. Another service is capable of finding property paths connecting two sets of resources.

The Jassa code base can be used both on the client and the server side. The only difference between these settings lies in the dependencies that need to be included. Furthermore, Jassa provides several RDF and SPARQL foundation classes which are nearly identical to those of the excellent API of the Java-based *Apache Jena*[6] project. The rationale followed by Jassa is to exploit existing, well-known API designs, rather than to invent new ones. At present, Jassa comprises the following modules:

- *util*: A utility module. Contains collections, such as *HashMap* and *HashSet*.
- *rdf*: The module that holds core classes related to RDF.
- *vocab*: A module for vocabularies, expressed in terms of classes of the *rdf* module.
- *sparql*: A module for core classes related to SPARQL. Builds upon the prior modules.
- *service*: Abstraction layer for SPARQL endpoints.
- *facete*: A faceted search module.
- *sponate*: A SPARQL-to-JSON mapper. Particularly powerful in combination with the generation of web frameworks that offer a clean separation between DOM and application logic, such as AngularJS and Ember.js.

The Jassa core classes aim to serve as a solid foundation for JavaScript-based Semantic Web applications. The *rdf* module provides the *Node* class (through *rdf.NodeFactory*) for encapsulating RDF terms. In contrast to approaches that are based on plain JSON, a Jassa node object provides methods such as *isLiteral()* and *getLiteralDatatypeUri()*. Furthermore, the *toString()* method is implemented to yield meaningful string representations. The *sparql* module contains several classes for the syntactic representation of SPARQL queries. An example showing a simple query for obtaining all DBpedia instances of type *Airport* is given in Listing 1. Note, that all mentioned namespaces reside in the global *jassa* object.

B. *Faceted geospatial browsing*

The Oxford Dictionary defines a *facet* as 'one side of something many-sided, especially of a cut gem'[7]. *Faceted classification* is an analytic-synthetic classification scheme which relies on multiple taxonomies to classify objects [5]. *Faceted browsing* builds upon faceted classification to enable navigation through information by applying multiple filters (i.e. *facets*). As far as Linked Data is concerned, tools such as *Sparklis*[8] and *Pelorus*[9] offer semantic faceted navigation over RDF datasets.

Jassa comes with a powerful SPARQL-based faceted search module in the *facete* namespace, which supports the definition of custom constraint types as well as constraining sets of resources by indirectly (possibly inversely) related properties. Unlike Sparklis and Pelorus, Jassa features support for nested and inverse properties and offers reusable components. Some recent development efforts which provide similar features are [6] and [7].

Faceted browsing widgets are implemented as AngularJS directives (Listing 2) and can thus be embedded into AngularJS applications using the corresponding HTML snippets. Note that the widgets are synchronized by AngularJS on the state of the *fctTreeConfig* object; any change will automatically trigger an update of the widgets.

Facete2[10] uses the above described features of Jassa and offers out of the box faceted browsing over SPARQL end-points to ease the navigation of RDF data using

```
1  var s = rdf.NodeFactory.createVar( 's' );
2  var o = rdf.NodeFactory.createUri( 'http :// dbpedia . org / ontology / Airport' );
3  var t = new rdf.Triple( s , vocab.rdf.type , o );
4
5  var query = new sparql.Query();
6  query.setResultStar( true );
7  query.setQueryPattern( new sparql.ElementTriplesBlock([ t ]) );
8  query.setLimit( 10 );
9  console.log( 'As string : ' + query );
10 // Output : Select * { ?s a < http :// dbpedia . org / ontology / Airport > } Limit 10
```

Listing 1. Forming a query with Jassa

[2] https://github.com/GeoKnow/Jassa-Core
[3] https://github.com/GeoKnow/Jassa-UI-Angular
[4] https://angularjs.org/
[5] https://github.com/AKSW/jena-sparql-api
[6] http://jena.apache.org
[7] http://www.oxforddictionaries.com/definition/english/facet
[8] http://www.irisa.fr/LIS/ferre/sparklis/
[9] http://clarkparsia.com/pelorus
[10] https://github.com/GeoKnow/Facete2

```
$scope.fctTreeConfig = new facete.FacetTreeConfig();
$scope.selectedPath = null; // Start with no selection
$scope.selectFacet = function( path ) { $scope.selectedPath = path; }
```

Listing 2. Showing values for a selected facet

advanced faceted search techniques. Spatial data is automatically detected and visualized on a map, even if the geometric information is only indirectly related to the resources specified by the faceted search.

Mappify[11] allows easy creation of embeddable map view snippets. It builds on reusable components of Facete2 and thus enables a facet based definition of points of interests based on a SPARQLaccessible dataset. Users are enabled to quickly style the map display by choosing marker icons and defining templates for the content to show when clicking the markers.

III. SPATIAL-SEMANTIC VISUALIZATION AND EXPLORATION ON MOBILE DEVICES

Our mobile spatial-semantic visualization and exploration tool is envisioned to give the user the ability to load the application with a custom dataset through one of the available SPARQL endpoints and retrieve the data they might be interested in on request, instead of (over)loading the mobile map with irrelevant information from the start.

Our goal is to deliver an easy to use, yet powerful mobile visualization and exploration tool that will provide a highly customizable and information-rich *slippy* map to the geospatial data consumers on the move. More specifically, the user interface should consist of:

- An interactive map component, to be used for quick and easy exploration of geographical areas, showing the user's own position (GPS coordinates) and the surrounding area;
- A semantic facet filtering component and a result view;
- A data source management component, to be used for quick loading and removing of visible information.

A. Slippy map

Leaflet.js[12] is an open-source JavaScript library for mobile-friendly interactive maps. It is extremely lightweight, yet has all the features we need for the task at hand. The library is designed with simplicity, performance and usability in mind. It works efficiently across all major desktop and mobile platforms, out of the box, taking advantage of HTML5 and CSS3 on modern browsers while still being accessible on older ones. It can also be extended with numerous plugins, and is an adequate lightweight replacement for the *OpenLayers*[13] library being used with Jassa for Facete2 and Mappify.

B. Platform

The identified requirements have narrowed down the choice of mobile development frameworks to *Apache Cordova*[14] / *Adobe Phonegap*[15] (formerly known *as Apache Callback*), which is an open-source set of device APIs that allow accessing native device function such as the camera or accelerometer from JavaScript. Combined with a UI framework such as *jQuery Mobile*[16] or *Dojo Mobile*[17] or *Sencha Touch*[18], this allows a smartphone app to be developed with just HTML, CSS, and JavaScript, which in turn, allows us to build upon some of the above mentioned existing work on desktop geospatial browsers, while focusing on the mobile user experience.

Since these JavaScript APIs are consistent across multiple device platforms and built on web standards, our solution is expected to be portable to other device platforms with minimal to no changes (Cordova is available for: iOS, Android, Blackberry, Windows Phone, Palm WebOS, Bada, and Symbian). This mobile development framework provides a set of uniform JavaScript libraries that can be invoked, with device-specific native backing code for those libraries. Moreover, the app would still be packaged using the platform SDKs and could be made available for installation from each device's application store (e.g. Google Play, Apple App Store, Windows Phone Store etc.).

IV. GEM

Our approach to the problem of putting the power of the above components in the hands of the mobile data consumers, titled GEM (Geospatial-semantic Exploration on the Move), relies on the previously mentioned, tailor-made and open-source Web technologies that are optimized for handheld devices.

The design philosophy behind GEM aims at maximizing the usable application/screen area through one stationary design component, and four on-demand widgets. The slippy map component represents the base layer on top of which the control widgets appear on request. The four components are given in Figures 1 and 2, and their functionalities described below:

- The *Facets* (left-hand) side drawer holds the loaded resource facet tree;
- The *Source Manager* (right-hand) side drawer is used to add/edit/remove available SPARQL endpoints / Linked Open Data sources;

[11] https://github.com/GeoKnow/Mappify
[12] http://leafletjs.com/
[13] http://openlayers.org/two/
[14] http://cordova.apache.org/
[15] http://phonegap.com/
[16] http://jquerymobile.com/
[17] http://dojotoolkit.org/features/mobile
[18] http://www.sencha.com/products/touch/

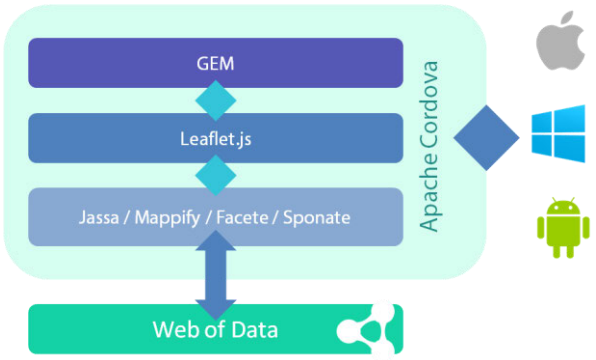


Figure 1. GEM technology stack

- The *Filter* text box (at the top), as the name suggests, is used to filter the resources on screen. The widget also holds the orientation indicator (i.e. compass);
- The *Resource details* bottom drawer pops-up to display the relevant information (e.g. label, URI and related triples) for the selected feature.

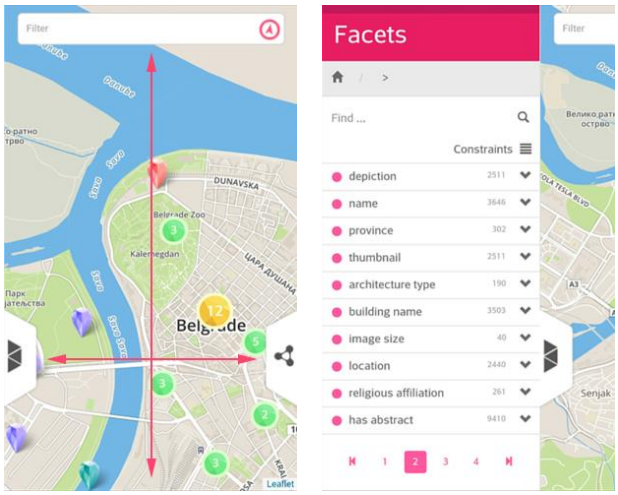


Figure 1. a) User interface; b) Facet tree side drawer

Different icon colors are used to indicate different resources / information sources. Moreover, to further prevent the map from overpopulating, and avoid overloading the mobile device resources in situations where multiple features are in each other's vicinity (relative to the map zoom level), we resort to marker clustering, i.e. grouping. Marker cluster groups are indicated by circles with resource counts. The color of the circle depends on the number of resources being grouped (ranging from green to red; green indicating small groups; Figure 1.a).

A. Data source management

The information shown on the screen is easily controlled using the Source Manager widget, which lets the user specify the name of the source (for convenience), the corresponding SPARQL endpoint and graph, as well as the desired resource type to be retrieved / displayed on the map. As Cordova relies on the host device's built-in Internet browser facilities, we exploit the browser's

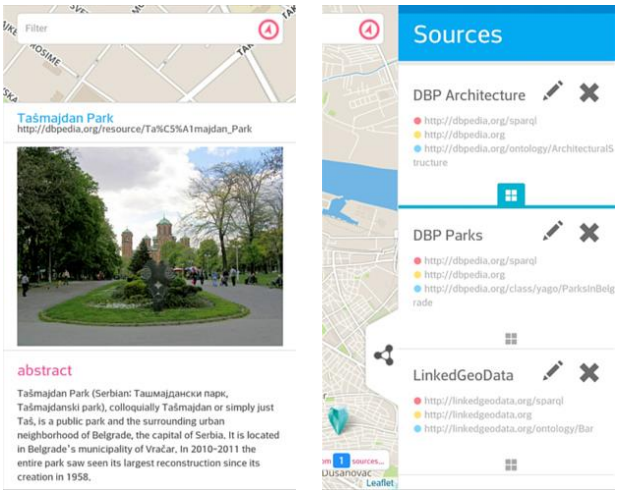


Figure 2. a) Resource details; b) Source management side drawer

HTML5 Local Storage to maintain the user preferences across sessions (i.e. automatically load the last used resource list the next time the application starts). Before HTML5, application data had to be stored in cookies, included in every server request. Local storage is more secure, and large amounts of various data can be stored locally, without affecting website/application performance.

B. Usage

We hereby outline the essential functionalities of the mobile geospatial-semantic browser provided by the above described facilities, and the basic steps to using them:

- The first time the application starts, GEM automatically populates the list of available LOD sources with DBpedia and LinkedGeoData SPARQL endpoints, using architectural sites, parks and bars as Base concepts (this is currently for demonstration purposes (e.g. a tourism-focused scenario) and might change in the future). The user can easily change this, by either editing the existing source list, or by creating a new one (i.e. by adding new endpoints / data types and/or removing the default ones; Figure 2.b).
- On startup, if the host OS location services are available (e.g. GPS), the map will zoom in on the user's own location (indicated by a gold pin), and the orientation indicator / compass in the filter widget will show (and update, based on) the direction the user's device is pointing to.
- Browsing the map will load the corresponding resources for the visible map area.
- To further make best use of the available screen space, upon detecting map interaction / movement, GEM will hide away the GUI elements the user cannot interact with while browsing through the map.
- Additionally, the application can also work in landscape mode.
- Should the user decide to narrow down their search, they can do so either by using the Filter box, or by using the Facet tree side drawer (Figure 1.b), which functions much in the way the Facete2 facet tree does.

The user first picks the facets (i.e. properties), and then the values to restrict the results to only those that fit the given description.

- Finally, to retrieve the details related to a single resource, the user only has to select the corresponding marker on the map. The bottom drawer reveals the feature's name (i.e. label) and URI. Clicking on the resource label expands the bottom drawer to display other relevant attributes (i.e. corresponding triples; Figure 2.a).

V. CONCLUSION

The GEM prototype described above represents a one-of-a-kind mobile faceted geospatial-semantic browser for Linked Open Data. It builds on top of the efforts already invested in state of the art open source desktop solutions and technologies, ranging from Jassa, Facete2 and Mappify to well-known third-party, community-maintained frameworks and libraries, such as AngularJS and Leaflet.js. It is deployed using Apache Cordova / Adobe Phonegap, allowing single-branch development for multiple major mobile platforms.

The current state of affairs opens up opportunities for future efforts, starting with backend improvements that will enable importing data based on other coordinate systems, allowing for richer and more complex data integrations. Moreover, an authoring component would give an entirely new dimension to the two-way interaction between the user and the application, and hopefully make a push towards wide community acceptance through crowdsourcing. Finally, further user experience improvements will be considered as well, such as automatic (re)source recommendations and tighter integration with Mappify, enabling easy sharing of information between the desktop and mobile clients, and making GEM an all-in-one solution for geospatial exploration on the move.

REFERENCES

[1] C. Becker and C. Bizer, *Exploring the Geospatial Semantic Web with DBpedia Mobile*. Journal of Web Semantics: Science, Services and Agents on the World Wide Web, vol. 7, pp. 278–286, 2009.

[2] C. J. Van Aart, B. J. Wielinga and W. R. van Hage, *Mobile Cultural Heritage Guide: Location-Aware Semantic Search*. In 17th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2010), vol. 6385 of Lecture Notes in Computer Science, pp. 257–271, 2010.

[3] M. Ruta, F. Scioscia, S. Ieva, G. Loseto, E. Di Sciascio, *Semantic Annotation of OpenStreetMap Points of Interest for Mobile Discovery and Navigation*. Proceedings of the 2012 IEEE First International Conference on Mobile Services, IEEE Computer Society, 2012.

[4] C. Stadler, P. Westphal and J. Lehmann, *Jassa - A JavaScript Suite for SPARQL-based Faceted Search*. Proceedings of the ISWC Developers Workshop 2014, co-located with the 13th International Semantic Web Conference, pp. 31–36, October 2014.

[5] W. Gödert, *Facet classification in online retrieval*. International Classification, Vol. 18 No. 2, pp. 98-109, 1991.

[6] M. Arenas, B. Cuenca Grau, E. Kharlamov, S. Marciuska, D. Zheleznyakov, and E. Jimenez-Ruiz, *Semfacet: Semantic faceted search over Yago*. In 23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, Companion Volume, pp. 123–126, 2014.

[7] H. Bast, F. Bäurle, B. Buchhold, and E. Haußmann, *Easy access to the freebase dataset*. In 23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, Companion Volume, pp. 95–98, 2014.

Cloud Network Infrastructure Design Approach

Vassil Gourov*, Elissaveta Gourova**, Borislav Lazarov*, Georgi Kostadinov*
* E-fellows Ltd., Sofia, Bulgaria
** Sofia University 'St. Kl. Ohridski', Sofia, Bulgaria
vgourov@efellows.bg
elis@fmi.uni-sofia.bg
blazarov@efellows.bg
gkostadinov@efellows.bg

***Abstract*—**During the past decade the cloud service market is one of the fastest growing segments around the world. The amount of companies that turned to the cloud has been steadily growing, since paying for a "shared" Cloud service over a given period of time reduces the capital expenditures and turned out to be better than using a dedicated hardware. This paper is focused on the architecture and the design of a shared public cloud service provider. The primary goal of the paper is to present a complete integrated solution for a single communication platform providing Cloud services to end-users. The paper, first, provides a literature review of some Cloud computing aspects, including the requirements for Cloud computing services and the key performance indicators to be evaluated. Second, it is described a real network infrastructure design approach, which allows smooth implementation of additional services and functionalities. All of the applicable functionalities are built to be managed and maintained separately for various independent customers in isolated mode of operations (Multi-Tenancy). The design allows optimal performance assuring high-availability of the service. Finally, the modular approach used in the design allows future optimisation and capacity upgrade plan of all key infrastructure components.

I. INTRODUCTION

Since the invention of the telephone more then a century ago rapid developments in all fields of science and technology have been witnessed. The trends in Information and Communication Technologies (ICT), especially, created enormous opportunities for fast access to data and information and their processing. While the development of telecommunications technologies facilitated the data transfer at high speed and regardless of geographical location, the Information Technologies (IT) created many new opportunities for increasing work efficiency of individuals, groups and organisations, for establishing new models of doing business, working, learning or entertaining.

The fast hardware developments following the Moore's law, on the one hand, and the increasing demands for computer power of the emerging applications, on the other, have forced organisations and individuals to regularly change their IT equipment. The heavy investments in technology, often underutilised [2], and the expected return of investments, as well as the increasing requirements for skills and knowledge for their deployment are among the driving forces in the uptake of Cloud computing [1]. This is closely related to a phenomenon that computers have become more powerful and less expensive, however, the pervasiveness of ICTs and the increasing management complexity is linked to growing expenses for organisations [2]. Subsequently, many organisations prefer nowadays to focus on their core capabilities and to outsource other functions introducing high overhead, such as ICTs. This is especially important for Small and Medium Enterprises (SME) which have difficulties to find highly-skilled professionals to maintain ICTs in-house [2], [4]. Thus, the opportunities to hire infrastructure or applications according to the real demands are decreasing the entry barriers for technology adoption in SMEs, and the time to market. Cloud computing offers also benefits like fast access to hardware resources without new capital investments, enhanced opportunities for scaling of services [2], as well as uninterrupted services and easier management [3]. At the same time, there are many evidences that the hardware emissions due to extensive IT use could be decreased by using Cloud services, contributing to more broad challenges of present-day economy like environmental sustainability [1], [3].

While the early providers of Cloud computing services are mainly in the United States [5], [7], there are evidences of their uptake also in Europe. Recently, a new Cloud infrastructure was developed in Bulgaria by eFellows Ltd. - a company established 10 years ago with a main focus on development and implementation of solutions in the field of information security, network and communication infrastructure, storage systems and data consolidation. Today the company has customers throughout the world in various fields.

The primary goal of the paper is to present the design approach followed by eFellows Ltd. for developing a Cloud-based network infrastructure. The first part of the paper provides an overview of Cloud computing and especially the concepts for Infrastructure as a Service (IaaS), and the challenges and requirements for its design. Second, a specific IaaS solution focused on meeting customer's needs is presented with special emphasis on physical and logical structure, and the orchestration options.

II. TRENDS IN INFRASTRUCTURE AS A SERVICE

A. Understanding of Cloud Computing

The Cloud computing phenomenon emerged in the 2000s, however, its core concepts could be traced back to the remotely connected terminals to mainframe computers, and later to the grid computing in academic institutions [1], [6]. Some authors [3] link this phenomenon to the advances in telecommunications, and in particular, the provision of Virtual Private Network

(VPN) services with comparable quality of service at a much lower cost. In fact, the emergence of Cloud computing was enabled by the uptake of three core technologies – Virtualisation, Multi-Tenancy and Web services [2]. By virtualisation the physical characteristics of the computing infrastructure or platform are abstracted and encapsulated, thus, hidden from end-users, and can be configured on demand, maintained and replicated very easily [2], [14]. In some cases these resources are partitioned (1:N) in multiple virtual elements, and in others – aggregated in a single virtual resource (N:1) [6]. In fact, through the virtualisation technique Cloud computing services are characterised by a front-end, seen by end-users, and a back-end, where the physical resources are configured, monitored and administrated by the providers [3]. This facilitates the Multi-Tenancy approach, whereas different users are served by the same resources without having any interference between them, and using independently the virtual resources allocated to them [2], [6]. Both, virtualisation and Multi-Tenancy, allow better utilisation of system resources available, thus leading to lower upfront and operational costs [2], [10], [14]. The Web-Services technology, on its side, facilitates the interoperability and interaction of different resources providing them standard interfaces over the network [2].

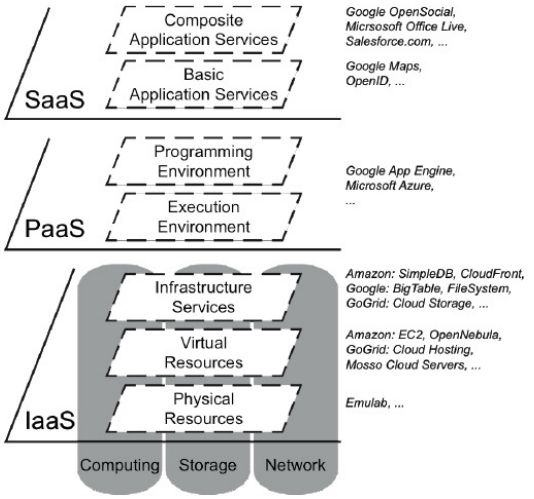


Figure 1. Layered Cloud computing infrastructure [6]

One of the widely quoted definitions of Cloud computing given by the National Institute of Standards and Technology [8] considers it as *"a model for enabling convenient, on-demand network access to a shared pool of configuration computing resources that can be rapidly provisioned and released with minimal management effort or service provider (SP) interaction."* It is widely accepted [3], [5], [6], [9], [11] that Cloud computing has three service layers (Fig. 1):

- *Software as a Service (SaaS)* ensures users access through a client network interface to commercially available software applications, which users do not need to install and maintain on their own computers.
- *Platform as a Service (PaaS)* ensures a computing environment where end-users can deploy their own software applications using the Cloud infrastructure of the provider, thus, avoiding additional costs for obtaining and supporting own computer platforms.
- *Infrastructure as a Service (IaaS)* provides end-users ICT infrastructure that is dynamically scalable on-demand, thus, allowing users to manage network and fundamental computing resources (e.g. servers and data centers) on which to run their operating systems and applications. IaaS providers manage the physical servers and network resources, and normally offer virtualised infrastructure as a service.

In comparison to computer grids, which are more research oriented, Cloud computing is more commercially oriented. Its basic characteristics comprise [2], [6], [14]: user friendliness and on-demand services, resource pooling, virtualised physical resources, architecture abstraction, dynamic scalability of resources, elastic and automated self-provisioning of resources, ubiquity, operational expense model, e.g. pay-per-use model, and Service Level Agreement (SLA). Specific for Cloud computing is the Multi-Tenancy, e.g. the provision of services to different types of users [9]: individual users, business users (large enterprises and governments, educational and research organisations and SMEs), developers and independent software vendors. The Cloud computing providers operate in two business models: variable (pay-for-your-usage) plans and fixed plans [5]. Variable plans allow customers to pay only for the resources actually consumed (e.g., instance hours, data transfer), and are consider as one of the major benefits of Cloud computing [9].

B. Requirements for Cloud services

The research literature [12] suggests that when designing an Internet-based system along with its functional characteristics, specific emphasis should be made on balancing its non-functional requirements: *availability* (depending on the reliability of all system components to work without any failure, and its robustness to cope with a failure), *performance* (measured with response time to users requests), *scalability* (ability to ensure performance when the number of users grows), *security* (ensuring normal system operation by controlling the access to its functionality while providing a reasonable degree of privacy), *manageability* (ability to monitor and alter the system runtime behavior), *maintainability* (how easy is to fix system problems or upgrade its components during runtime), *flexibility* (ability to produce new system versions or re-configure it) *and portability* (easy migration to a new environment). In the case of Cloud computing, most of these requirements are stressed by researchers as well [1], [6], [9], [13]. For example, Venters and Whitley [1] provide an overview of the specific requirements to Cloud computing systems (Table I). It is interesting to note their emphasis on providing equivalent opportunities to Cloud users in terms of security, response time and performance compared to similar services, and at the same time, offering benefits linked to service variety and scalability. In addition to these requirements, Develder et al. [6] consider the specific needs of different applications (targeted at scientific, business or individual users) in terms of resource volume and granularity (e.g. storage volumes, CPU performance and network bandwidth), elasticity and multiple tasks opportunities. As a specific issue is pointed out also the ability of Cloud services to be integrated with those available in-house [6].

TABLE I.
REQUIREMENTS TO CLOUD COMPUTING, ADAPTED [1]

technological dimensions	
Security Equivalence	at least equivalent in security to that experienced when using a locally running server
Availability Equivalence	at least equivalent in availability to that experienced when using a local server
Latency Equivalence	at least equivalent in latency to that experienced when using a locally running server
Variety	provides variety corresponding with the use for which the service will be put
Abstraction	abstract away unnecessary complexity for the service they provide
Scalability	service which is scalable to meet demand
service dimensions	
Efficiency	helps users be more efficient economically
Creativity	aids innovation and creativity
Simplicity	simple to understand and use

Garg et al. [13] point out that the Virtual Machine (VM) performance often varies from the promised values in the SLA, which reflects on the Quality of Services (QaS) offered to clients. Therefore, the authors propose a Service Measurement Index (SMI) comprising a set of indicators which could be used for taking a decision which Cloud SP better meets the specific user's requirements. In particular, for assessment of IaaS providers the authors suggest the following Key Performance Indicators (KPI):

- *Service response time* - measured in terms of the response time for making the service available for usage. In the case of IaaS, this includes provisioning the VM, booting the VM, assigning an IP address and starting application deployment;
- *Sustainability* - defined in terms of the environmental impact, e.g. can be measured as the average carbon footprint or energy efficiency of the Cloud service;
- *Suitability* - the degree to which users requirements are met, including both functional and non-functional requirements;
- *Accuracy* - measures the degree of proximity to the user's actual values when using a service compared to the expected values given in the SLA, e.g. the frequency of failure in fulfilling the promised SLA in terms of Compute units, network, and storage;
- *Transparency* - indicates the extent to which usability is affected by any changes in service, and can be measured as a time for which the performance of the user's application is affected during a change in the service or the frequency of such effects;
- *Interoperability* - ability of a service to interact with other services offered 0by the same provider or other);
- *Availability* - percentage of time a customer can access the service;
- *Reliability* - reflects how a service operates without failure during a given time and conditions, based on the mean time to failure promised by the provider and previous failures experienced by the users;
- *Cost* - depends on service acquisition and usage, e.g. could be measured according to the cost of one unit of CPU, storage, RAM, and network bandwidth;
- *Adaptability* - ability of the Cloud provider to adjust changes in services based on users requests;
- *Elasticity* - defined in terms of how much a Cloud service can be scaled during peak times;
- *Usability* - the ease of use could be measured in terms of the average time experienced by the previous users of the Cloud service to operate, learn, install and understand it;
- *Throughput* - evaluate the performance of infrastructure services, and depends on several factors that can affect execution of a task, e.g. number of tasks of the user application and the number of machines on which it runs;
- *System efficiency* - indicates the effective utilisation of leased services, e.g. its higher value is linked to a smaller overhead;
- *Scalability* - determines whether a system can handle a large number of application requests simultaneously, e.g. the ability to scale resources horizontally (e.g. 'scale out' Cloud resources of the same types) or vertically ('scale up' different Cloud resources assigned to a particular Cloud service).

In order to meet users' requirements, and more specifically, ensure guaranteed QoS and meet the provisions of the SLA, it is essential to undertake special monitoring and management efforts. As stressed by Mohamaddiah et al. [11], resource management in Cloud computing should focus on three interrelated processes: monitoring (infrastructure management and control using KPIs), allocation (assigning available recourses) and discovery (determining the most appropriate resources to meet SLA). Generally, the Services providers assign specific resources to the incoming job requests from end-users. This requires real-time information on the load and availability of physical resources. Therefore, the Infrastructure Provider monitors and controls the infrastructure performance and takes care of its optimal utilisation. In case of alack of physical resources, these could be hired from other IP or in case of higher availability – offered to them [11]. The management of Cloud resources is very important not only for the services provided to end-users, but also for the interrelations among Cloud providers in federated Cloud structures mediated by brokers [10].

III. DEVELOPMENT OF CLOUD INFRASTRUCTURE

A. Background

The main goal is to design a complete integrated solution for a single communication platform providing Cloud services to end users and internal users (company employees). The network infrastructure design should allow smooth implementation of additional services and functionalities. All of the applicable functionalities should ensure separate management and maintenance for different independent customers in isolated mode of operations (Multi-Tenancy), meeting all the Cloud services providers' special requirements. Some functionalities should be automatically activated and configured by the end users through a Self-Service Portal. All self-service features should be realised through a custom built "Orchestration" application, developed to automatically execute predefined template scripts (workflows) for configuration of specific functionalities and settings. Orchestration application should be tightly

integrated with the self-service portal to be also custom built in-house.

Other requirements for the infrastructure solution comprise:

- the logical and the physical solution designs should allow optimal performance, graceful degradation (fault tolerance) and increased availability and capabilities of the devices and interconnections between them;
- the solution should be built with redundancy at minimum 2:1, so that no Single Points of Failure components to be allowed;
- the solution should ensure scalability, e.g. to ensure future optimisation and capacity upgrade opportunities of all key infrastructure modules due to the need of growth, increasing number of customers/users or other system requirements.

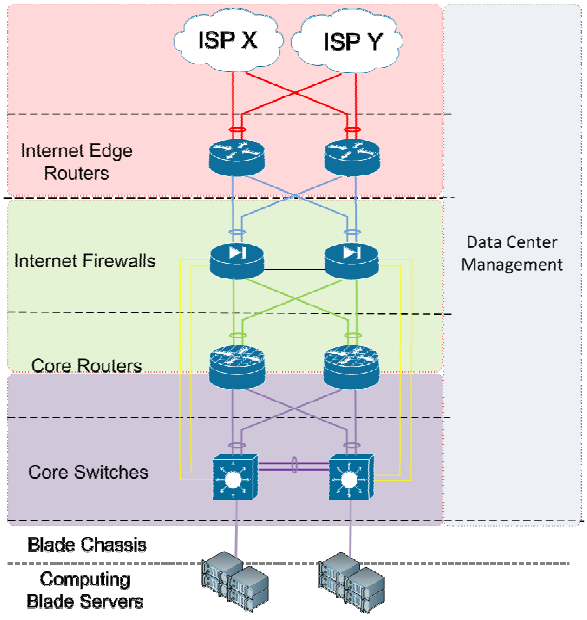


Figure 2. Simplified Physical Network Diagram

B. *Concept for IaaS*

In order to ensure these requirements, during the design phase were taken into consideration the best world and industry practices and manufacturers' recommendations. Special emphasis was put on the physical and logical network design. In order to meet the requirements for high performance, flexibility, scalability and high availability of the entire network infrastructure, it was separated into layers, including (Fig. 2):

- *Blade Chassis Networking:* consisting of some blade servers and converged network modules, whereas the latter provide additional layer of abstraction by presenting to the blade servers virtual Network Interface Controllers (NIC) over the physical 10 gbps NICs for both data and storage. They also provide external network connectivity over 1G/10G Ethernet and FibreChannel.
- *Data Center Switches:* The Core Switches have a function to provide switching between all physical and virtual hosts within the Cloud computing environment. A Data Center Access and management Switch: providing network access for single servers (outside the blade chassis) as well as Out-of-Band (OOB) management interfaces of any Data Center equipment (servers, storage, network).
- *Data Center Core Routers:* used to provide routing for all the subnets within the datacenter (both for clients/tenants and for internal/system use), as well as to provide VPN termination for Site-to-Site and Remote Access VPNs.
- *Internet Firewalls:* including devices that perform Network Address Translation (NAT) between private Virtual Machines (VM) IP addresses and public IP addresses, ensure Internet access control and security policies.
- *Wide-Area Network (WAN) Switches:* providing any-to-any full mesh connectivity between all WAN links (such as Internet Service Providers (ISPs), Metropolitan Area Network (MAN) transport, etc.) and the corresponding termination devices (Internet Edge Routers, Data Center Core Routers, other). In addition, these switches perform also the function of Edge Interconnect Switches in order to achieve resource optimisation.
- *Internet Edge Routers:* connecting to Upstream ISPs and exchange Border Gateway Protocol (BGP) routing information, and guaranteeing, thus, outbound and inbound network reachability for the Provider Independent IP network and company Internet Autonomous System.

The Logical network design (Fig. 3) determines how the entire network is divided into individual segments and how packets are routed between each of these segments. Generally, the network is divided into zones containing one or more individual networks with similar functions, connectivity and security needs. For the customers the internet firewalls split the cloud network into four security zones. The outside zone is the connection with the Internet. There are two demilitarised zones (DMZ) namely public and private. In the public zone all publicly available common and shared resources that need to be access from external sources over the Internet are placed here – e.g. Public DNS servers. The private DMZ is where all common and shared cloud resources, that need to be accessed by all or most of the internal client VMs are be placed. There are one or more subnets with specific access policies assigned to this zone and they are used for services like DHCP, DNS and etc. Furthermore, there is the inside zone, where the client networks reside. . Each of those networks represent a separate 802.1Q VLAN and are also reside in a different L3 routing table. This is how complete tenant isolation is archived. Additionally, there is a dedicated transport network between the core routers and the firewalls. All Client VM traffic destined to the Internet or to the Public/Private DMZ resources will be routed via this network. iBGP dynamic routing protocol is used to provide network exchange information and to ensure high availability and load balancing between the devices. Finally, there is a specialised zone where all remote access VPN (Similar to Dial-UP) users will be terminated and will have their VPN IP Address allocated. In the design five VPN client use cases are considered – office Local Area Networks (LAN), private Cloud, home LAN, mobile worker and home worker.

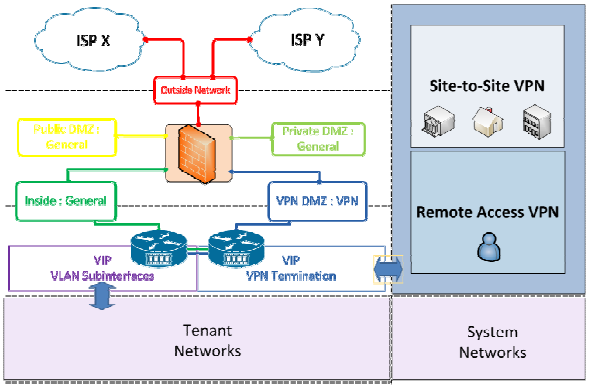


Figure 3. Logical Network Diagram

This design supports two VPN types: Site-to-site VPN and Remote Access VPN. The former provides permanent secure connectivity between the client's resources within the Cloud and a remote location based on infrastructure device(s) such as router or firewall. The Remote Access VPN type provides on-demand secure access to client resources within the Cloud via the Internet similar to dial-up, whereas users authenticate with username and password, configured via the self-service portal.

Internet connectivity is ensured by registering own Autonomous System and public IP Address ranges. BGP peering with any ISP can be established using them to guarantee adequate performance and high availability for the users. The design and the equipment supports adding new connections without service disruption.

All client Virtual Machines are allowed for outbound Internet connectivity using NAT performed by the Firewall by default. No additional configuration is required. Special features such as Anti-Virus and Web Content Filtering are supported by the Firewall and can be configured manually on demand.

Inbound Internet connectivity for client virtual machine is possible using static NAT (Virtual IPs) performed by the firewall. It can be configured automatically by the orchestration service on user requests in the self-service portal. Users are able to activate static or dynamic public IPs bound to a particular internal IP address of a virtual machine in the Cloud and to apply security filters based on IP protocol and port.

Management and Monitoring tools ensure:

- *Log Collection:* All infrastructure devices are configured to export event logs to an external server using standard SYSLOG protocol, which is useful both in routine troubleshooting and in incident handling.
- *Configuration Management and Backup:* All infrastructure devices are configured to archive configuration changes and to automatically backup configuration.
- *Monitoring:* All vital infrastructure components are being continuously monitored and graphed. If any component fails or falls out of acceptable thresholds an automatic notification is being sent to the support team.
- *Management Access to equipment:* Infrastructure devices' management interfaces require authentication with valid credentials. Only valid users who are members of a specific domain group are entitled with management access. Furthermore, there are access lists set on every device that allow access only from certain IP address ranges – used by the support team.

All backend operations and configurations that should be performed on the infrastructure triggered by end-user actions via self-service portal. The Orchestration options in case of Physical Networking are depicted in Fig. 4. All self-service features are realised through a custom built "Orchestration" application, developed to automatically execute predefined template scripts (workflows) for configuration of specific functionalities and setting. The communication between the application and the infrastructure is archived via standard command line interfaces (CLI) such as SSH and PowerShell.

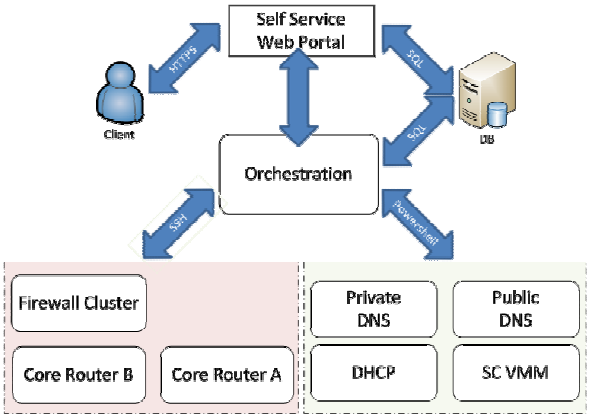


Figure 4. Physical Network Orchestration workflow

C. *Main features of the approach*

Usability: The self-service portal is easy-to-use and allows the users to build the server they want right away. From there on the clients can install and configure any software they want.

Reliability: Both, the logical and the physical solution designs allow optimal performance, graceful degradation (fault tolerance) and increased availability and capabilities of the devices and interconnections between them. Also, the solution has been built with redundancy at minimum 2:1, so that there are no Single Points of Failure components.

Scalability: The design includes future optimisation and capacity upgrade plan of all key infrastructure modules (components based or overall solution oriented) due to the need of growth, increasing number of customers/users or other system requirements.

Performance: All components of the infrastructure are in Active/Active state. This means that the traffic from the VMs is balanced trough one of the device pairs based on network segment the traffic comes from. The network throughput is 20 Gbps for inter-VM traffic, 5 Gbps for L3 interconnections and 600 Mbps for the Internet links.

IV. CONCLUSIONS

The paper provides an overview of the current cloud service requirements and evaluation. It gives an architectural description of a newly deployed cloud service provider in Bulgaria. The main features of the

suggested design are High Availability, Load Balancing and Multi-Tenant Isolation. Furthermore, the architecture uses a modular approach so the components can be easily upgraded.

The main problem of the approach is that it support a limited number of tenants. It is possible to add more devices horizontally, but this is not a very cost effective solution. This has been taken into consideration

REFERENCES

[1] W. Venters, E. A. Whitley, A critical review of cloud computing: researching desires and realities, *Journal of Information Technology*, 27, 2012, pp. 179–197.

[2] S. Marston, Z. Li, S. Bandyopadhyay, J. Zhang, A. Ghalsasi, Cloud computing — The business perspective, *Decision Support Systems*, 51, 2011, pp. 176–189.

[3] Y. Jadeja, K. Modi, Cloud Computing - Concepts, Architecture and Challenges, *International Conference on Computing, Electronics and Electrical Technologies*, 2012, pp. 877-880.

[4] E. Gourova, V. Kadrev, A. Stancheva, G. K. Petrov, M. Dragomirova, Adapting educational programmes according to e-competence needs: the Bulgarian case", *Interactive Technology and Smart Education*, 11(2), 2014, pp. 123-145.

[5] Y. Han, Cloud Computing: Case Studies and Total Costs of Ownership, *Information Technology and Libraries*, *30*(4), 2011, pp. 198-206.

[6] Ch. Develder, M. De Leenheer, B. Dhoedt, M. Pickavet, D. Colle, P. Demeester, Optical Networks for Grid and Cloud Computing Applications, *Proceedings of the IEEE*, *100*(5), 2012, pp. 1149-1167.

[7] Schubert L. *The future of Cloud Computing: Opportunities for European Cloud Computing beyond 2010*. Exper Group Report: public version 1.0. European Commission, 2011.

[8] P. Mell, T. Grance, *The NIST Definition of Cloud Computing*, NIST, http://csrc.nist.gov/groups/SNS/cloud computing/ (accessed Oct. 21, 2010).

[9] S. Patidar, D. Rane, P. Jain, A Survey Paper on Cloud Computing, *Second International Conference on Advanced Computing & Communication Technologies*, 2012, pp. 394-397.

[10] D. Villegas, N. Bobroff, I. Rodero, J. Delgado, Y. Liu, A. Devarakonda, L. Fong, S. M. Sadjadi, M. Parashar, Cloud federation in a layered service model, *Journal of Computer and System Sciences*, 78, 2012, pp. 1330–1344.

[11] M. H. Mohamaddiah, A. Abdullah, M. Hussin, S. Subramaniam,A Proposed Architectural Framework for Resource Provisioning Mechanism in Cloud Computing, *1st International Conference of Recent Trends in Information and Communication Technologies*, Sept. 2014, pp. 312-327.

[12] P. Dyson, A. Longshaw, *Architecting Enterprise Solutions: Patterns for High-Capability Internet-Based Systems*, John Wiley & Sons, 2004.

[13] S. K. Garg, S. Versteeg, R. Buyya, A framework for ranking of cloud computing services, *Future Generation Computer Systems*, 29, 2013, pp. 1012–1023.

[14] U. Divakarla, G. Kumari, An Overview Of Cloud Computing In Distributed Systems, *AIP Conference Proceedings*, 1324(1), 2010, pp. 184-186.

A Routing Algorithm for Mobile Ad Hoc Networks

Ivan Djokic, Aldina Avdic, Aleksandra Pavlovic
State University of Novi Pazar / Department of Technical Sciences, Novi Pazar, Serbia
idjokic@np.ac.rs, apljaskovic@np.ac.rs, apavlovic@np.ac.rs

***Abstract*—A mobile ad hoc network is a set of mobile nodes that are dynamically located and connected by wireless links. The network is self-configuring and provides end-to-end communication. In order to facilitate communication within the network, a routing protocol is used to discover routes between source and destination nodes. Routing in the mobile ad hoc networks is a challenging task and has received a tremendous amount of attention from researches. This has led to development of many different routing protocols. Therefore, it is quite difficult to determine which protocols may perform best under a number of different network scenarios, such as increasing node density and traffic. In this paper, we present concept, characteristics and functionality of a simple routing protocol, based on packet delivery rate and distance from the destination node.**

I. INTRODUCTION

Wireless mobile ad-hoc networks have no fixed infrastructure. A dynamic routing protocol is needed to function properly on a frequently changing network topology. Here the node itself acts as both client and server, forwarding and receiving packets to or from other nodes. Routing in ad-hoc networks has become a challenging issue. There are many protocols already developed for mobile network environments. All these protocols can be classified in different ways. Based on the network structure the routing protocols can be classified as flat routing, hierarchical routing and geographic position assisted routing [1]. In flat routing, nodes communicate directly with each other. The flat routing protocols can be classified in three categories such as proactive, reactive and hybrid. Proactive protocols follow the strategies which are mostly followed by conventional routing protocols. On-demand routing is a new emerging technology in ad-hoc networks. Hybrid protocols are incorporating the properties of both proactive and reactive types. Hierarchical routing plays a major role in large size networks where flat routing protocols are struggling with constraints. Now-a-days geographical location information also provides better routing performance in ad-hoc networks.

In proactive scheme, a very small delay is needed to determine the route but a significant amount of delay is needed for creating a route by reactive routing protocols. Pure proactive scheme is not appropriate for the ad-hoc networking environment, because it has to keep the current routing information in a large network. Reactive protocols require significant control traffic due to the long delay and excessive control traffic. As a result pure reactive routing protocols are not suitable for large network implementations.

The focus of this paper is the concept of a new ad hoc network routing protocol based on following data: (1) the source - neighboring nodes packet delivery rate, and (2) the distance between destination and source neighboring nodes.

II. RELATED WORK

The history of wireless networks started in the 1970s and the interest has been growing ever since. A new generation is the construction of temporary networks with no wires, no communication infrastructure and no administrative intervention required. Such interconnection between mobile computers is called an ad hoc network. Ad hoc networks are defined as a collection of mobile nodes forming a temporary (spontaneous) network without the aid of any centralized administration or standard support services. [2]. Routing protocols for Mobile ad hoc networks can be broadly classified into two main categories: (1) proactive or table-driven routing protocols and (2) reactive or on-demand routing protocols. In proactive or table-driven routing protocols, each node continuously maintains up-to-date routes to every other node in the network. Routing information is periodically transmitted throughout the network in order to maintain routing table consistency. Certain proactive routing protocols are Destination- Sequenced Distance Vector (DSDV), Wireless Routing Protocol (WRP), Global State Routing (GSR) and Clusterhead Gateway Switch Routing (CGSR). In contrast to proactive approach, in reactive or on demand protocols, a node initiates a route discovery throughout the network, only when it wants to send packets to its destination. Some reactive protocols are Cluster Based Routing Protocol (CBRP), Ad hoc On-Demand Distance Vector (AODV), Dynamic Source Routing (DSR), Temporally Ordered Routing Algorithm (TORA), Associatively-Based Routing (ABR), Signal Stability Routing (SSR) and Location Aided Routing (LAR) [3]. Till now many routing protocols are used in mobile ad hoc networks. Each routing protocol has unique features. Based on network environments, we have to choose the suitable routing protocol. Protocol behavioral study is conducted very intensively. Proactive routing protocols are best suited in small networks. In large and dense network, proactive routing protocols cannot perform well. Proactive routing protocols are table driven. Maintaining thousands of routing tables properly in large network degrades the efficiency. So for large and dense networks reactive routing approach plays a major role. Reactive routing protocols use destination sequence number and feasible distance to ensure a loop free routing. Hybrid routing protocols use reactive and proactive approach in routing operations [4]. Some works consider

the routing task, in a way that a message is to be sent from a source node to a destination node, where the destination node is known and addressed by means of its location. Routing is performed by a scheme based on this information, generally classified as a position-based scheme [5]. There are many challenges to be faced in routing protocols design. A central challenge is the development of the dynamic routing protocol that can efficiently find routes between two communication nodes. Also, in order to analyze and improve existing or new MANET routing protocols, it is desirable to examine other metrics like power consumption, fault tolerance, number of hops, jitter, etc. in various mobility and traffic models [6].

III. THE PROPOSED NEW ROUTING PROTOCOL

Here we present a new on-demand, localized, and packet delivery rate based Ad Hoc routing protocol. Localized algorithms are distributed in nature and resemble greedy algorithms, where simple local behavior achieves a desired global objective. In a localized routing algorithm, each node makes a decision to which neighbor to forward the message based solely on the location of itself, its neighboring nodes, and the destination. On the other hand, the primary goal of every routing scheme is to deliver the message, and the best assurance one can offer is to design a routing scheme that will guarantee delivery. Wireless networks normally use a single-frequency communication model where a message intended for a neighbor is heard by all other neighbors within the transmission radius of the sender. Collisions normally occur in medium access schemes, such as IEEE 802.11. This protocol is based on guaranteed delivery in routing (i.e., eventual delivery), which is conditional on the ability of the medium access layer to always transmit a message between any two neighboring nodes, possibly with retransmissions.

In proposed routing protocol, a routing is constructed only when a node needs to communicate with another node. Assume that a source node S wants to send a packet to some destination node D. Then, the routing algorithm follows the next nine rules:

1. The algorithm is running on the S node, when it has a data packet for delivering to D node, where S and D are nodes of an ad hoc network;
2. The algorithm input data are address and location of D node, and the minimum packet delivery rate between nodes on the route from S to D (PDRmin);
3. The S node determines the neighboring nodes (nodes within the range);
4. The S node determines or measures packet delivery rate to the neighboring nodes (with 10 short ping messages, for example);
5. The S node selects the neighboring nodes with PDR≥PDRmin;
6. The S node obtains locations of neighboring nodes with PDR≥PDRmin;
7. The S node calculates distance from nodes with PDR≥PDRmin to the D node;
8. The S node selects the node with the shortest distance to the D node (next hop node);
9. The S node starts the data transmission to the next hop node.

Figure 1 shows selection of the next hop node by proposed routing algorithm, with PDRmin=90%. The next hop node has PDRmin≥90% and the distance from D node R1min.

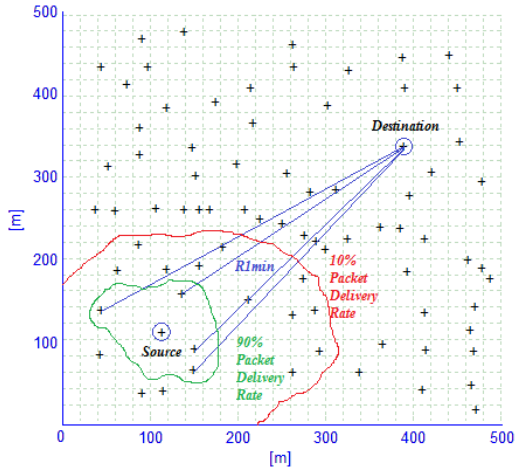


Figure 1. Next hop node selection

Now, the next hop node becomes S node and it applies the same routing algorithm. The algorithm proceeds until the destination is reached or no closer node to the destination exists. Varying PDRmin different nodes will be selected to form S-D route. Figure 2 shows two routes with two different packet delivery rates between two consecutive nodes, PDRmin≥0.9 and PDRmin≥0.75. Obviously, a lower packet delivery rate gives a route with fewer hops. However, this does not mean better efficiency. Generally, we can conclude that in low mobility and low load scenarios our proposed protocol finds S-D route for different network topologies and different packet delivery rates between nodes. But, there are many other challenges to be faced in the proposed protocol evaluation and implementation.

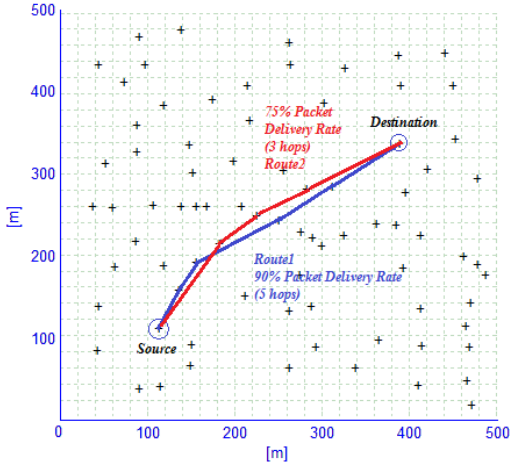


Figure 2. Network routes with different packet delivery rates

In order to analyze and improve this new mobile network routing protocol, it is desirable to examine other metrics like node mobility, power consumption, fault tolerance, jitter, etc. in various traffic models. Preliminary results show robustness of the proposed algorithm, it handles the position deviation due to the dynamicity of the network, and has the ability to deliver a message when the communication model deviates from the unit graph, due to obstacles or noise.

IV. PROTOCOL SIMULATION

In order to present which are the results of the proposed routing algorithm in practice, a small simulation application is made using programming environment MS Visual C # 2010 Express. The basis for the program was a pseudo code shown in Fig 3, which was written according to the routing algorithm given in the previous section.

```
procedure RoutingProtocol(Destination, Package, PDRmin)
NextHop = Source;
while NextHop ≠ Destination do
begin
foreach Node ∈ AdHocNetwork
  begin
        PDR[NodeId]
        = FindPDRofNode(NextHop, PingMessage, Node);

        If (Node.PDR>=PDRMin)
              AddNodeToRangeList(Node);
  end
foreach Node ∈ RangeList
  begin
        Distance[NodeId]
        = FindDistanceFromDestination(Destination);

        MinimumDistanceNode
        = FindMinimumDistanceNode(Distance[NodeId]);

        sendPackage(NextHop, MinimumDistanceNode);
        NextHop = MinimumDistanceNode;
        ClearRangeList();
  end
end
```

Figure 3. Routing algorithm pseudo code

The appearance of the application for the simulation, after running of an routing example, is shown in Fig 4. Input data are the number of nodes in ad hoc networks, the source position, the destination position and minimum PDR. After that, nodes are deployed randomly in the diagram, and the PDR for each of them is assigned based on distance. In conditions of applying the algorithm in the real network, the PDR will be determined by sending a certain number of short ping messages.

The application operates as follows. After entering the position of the source and the destination in given text boxes, those positions are plotted on the diagram, and after the election of the PDRmin, the diagram is marked with a green area which consists of nodes that satisfy the condition that the PDR >= PDRmin (range). Then, the next hop (the node to which the packet will be forwarded) is selected as node from the range nearest to the destination. It will be, in the next step of simulation marked as a new source (a black triangle symbol). In next step of the simulation, the diagram shows the range of the new source, and the next hop is marked with symbol x black. This procedure will be repeated until the packet reaches the destination.

Based on this simulation, it is shown that the package will be delivered to the destination, and which route will be selected, and the number of nodes that package will pass through, depending on the area of the range, that is the determined by value PDRmin. All this is recorded in the log and report in the application for the routing algorithm simulation.

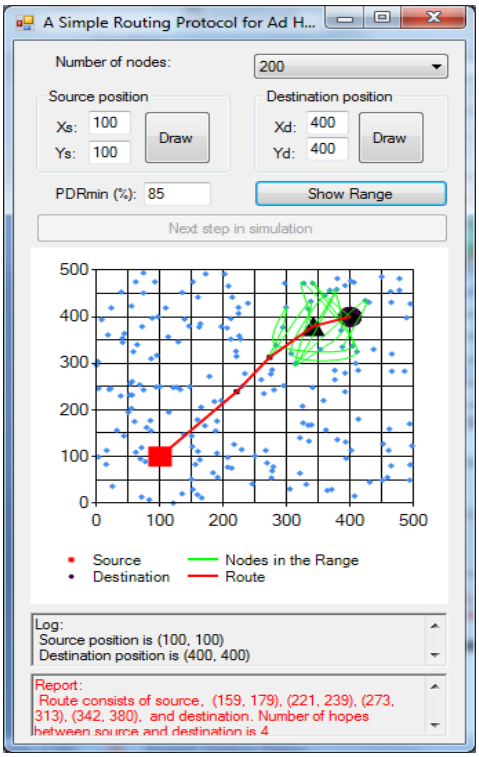


Figure 4. Application for the algorithm simulation

The example shows the route found using the routing protocol in condition when the source position is (100, 100), the destination position is (400, 400), and PDRmin is 85%, and other nodes in an ad hoc network are placed as it is shown in Fig 4.

V. CONCLUSIONS

In this paper, we present a new on-demand routing protocol based on the packet delivery rate between source and neighboring nodes, and the distance between source neighboring nodes and the destination node. Preliminary results show that the proposed protocol algorithm is able to quickly find a source destination route for different network topologies and different required packet delivery rates between nodes. The algorithm also exhibit the properties of robustness, it has the ability to deliver a message when the communication model deviates from the unit graph due to obstacles or noise, but influence of node mobility has to be evaluated through different scenarios.

ACKNOWLEDGMENT

This research was partially supported by Ministry of Science of Serbia, under the grant TR32023, TR35026 and III44007.

REFERENCES

[1] X. Hong, K. Xu, and M. Gerla, "Scalable Routing Protocols for Mobile Ad-Hoc Networks" *IEEE Network Magazine*, vol. 16, pp. 11–21, 2002.

[2] A. K. Gupta, H. Sadawarti and A. K. Verma. "Performance analysis of AODV, DSR & TORA Routing Protocols" *IACSIT International Journal of Engineering and Technology*, vol.2, April 2010.

[3] V. Kumar and R. Sharma "Behavioral Study of Dynamic Routing Protocols for MANET", *International Journal of Computing and Business Research*, vol. 2, May 2011.

[4] J. Uddin and M. R. Zasad, "Study and Performance Comparison of MANET Routing Protocols: TORA, LDR and ZRP", Department of Electrical Engineering School of Engineering, Blekinge Institute of Technology, Karlskrona, Sweden, 2010.

[5] I. Stojmenovic, "Position-Based Routing in Ad Hoc Networks", *IEEE Communications Magazine*, vol. 40, pp. 128-134, July 2002.

[6] S. Baraković, S. Kasapović, and J. Baraković, "Comparison of MANET Routing Protocols in Different Traffic and Mobility Models", *Telfor Journal*, vol. 2, pp. 8-13, 2010.

[7] A. Boukerche, *Algorithms and Protocols for Wireless and Mobile ad hoc networks*, Wiley, 2009.

[8] A. Vasilakos, Y. Zhang, and T. V. Spyropoulos, *Delay Tolerant Networks – Protocols and Aplications*, CRC Press, 2011.

[9] M.Y. Rhee, *Wireless Mobile Internet Security*, Wiley, 2013.

[10] H. Zhang, S. Olariu, J. Cao and D. Johnson, "Mobile Ad-hoc and Sensor Networks", *Third International Conference, Beijing*, 2007.

[11] S. Misra, I. Woungang and S. C. Misra, *Guide to Wireless Ad Hoc Networks*, Springer, 2009.

[12] H. Shuang, *Multicast Routing Protocols in Mobile Ad Hoc Networks*, ProQuest LLC, 2009.

[13] R. Graziani and A. Johnson, *Routing Protocols and Concepts*, Cisco Systems, 2008.

Linked data network approach to ontology-based data sharing

Igor Miletic*, Zoran Marjanovic**, Miroslav Ljubicic**
*Research consultant, Belgrade, Serbia
**Faculty of Organizational Sciences, Belgrade, Serbia
igor.miletic@linuxmail.org, marjanovic.zoran@fon.rs, ljubicic.miroslav@fon.rs

***Abstract*—Enterprise systems, such as Enterprise Resource Planning systems, usually keep their data isolated and not readily available to the other systems. When an enterprise system needs to open data for external usage, it is usually done using interface tables, Web services, or some kind of data exports. These approaches require both the source system, which provides data, and the destination system, which consumes data, to implement additional routines in order to support data sharing. Our approach uses linked data networks and introduces ontologies as a means to define standard data structures that enterprise systems can use to expose their data. This way, the exposed data, supported by auxiliary tools, can be referenced, read or updated from various systems in a common and unified way by collaboration systems. Our approach allows this data to become a part of much bigger linked data network.**

I. INTRODUCTION

Today's industry approaches to data sharing require significant, additional work for every new situation where source and destination systems need to interface. Enterprise systems, such as Enterprise Resource Planning (ERP) systems, usually keep their data isolated and not available to the other systems. Of course, this is an understandable decision that follows from the concerns for privacy and security. Typically, then, when an enterprise system needs to open data for external usage, it is done using interface tables, Web services or some other kind of data export mechanism. These approaches require both the source system, which provides data, and the destination system, which consumes data, to implement additional routines in order to support data sharing. By data sharing we mean disclosure of data from the source system and access of data by the destination system. Mainly, we are focusing on data sharing between business systems, such as ERP systems, because these types of systems are traditionally treated as isolated systems.

These industry approaches towards data sharing are essentially designed and implemented without taking into account longer-term or broader collaboration needs of the enterprise at the application domain level. Sharing data between two or more systems mostly relies on software components (services) that provide data using one of most common formats, such as XML[1], JSON[2], CSV[3], and so on. Rarely, data sharing is provided via database or some kind of persistent format like files via FTP[4] or similar. Step forward in data sharing are data exchange standards (e.g. EDI, RosettaNet, ebXML), which provide guidance how exchange message should look like. The fact is that today we don't have precise rules or guidance that one system should follow when sharing data. Sharing solution always depends on a specific circumstance, programming language, protocols or data to be shared.

Yet, the result of such ad hoc approaches is a significant cost of data sharing and data redundancy incurred over the life-time of the enterprise. This is manifested by significant, incremental costs of these ad hoc integrations. We observe that the biggest effort and high costs, related to data sharing, exist because every single system needs to convert and store data in their local data structure. Even data that are natively occurring in external systems need to be converted and persisted in a local data structure because systems are usually designed to work only with their local data. There is always the need for extra effort to convert data in the sense of their structure, data types, constraints, etc. Beside the costs, these conversions are also generators of data redundancy, as we are forced to store same data in two or more different structures. As number of integrated systems grows, redundancy grows, too.

Ontology-based approaches to data sharing have been explored extensively in the past decade with promising results. Researchers have investigated use of ontologies as standard application domain data structures that enterprise systems may use to expose their data [4, 6, 7]. Results of the research suggest that ontologies may be used as very good base-line for data-centric systems which will have good outcomes (i.e., less effort to integrate, lower costs, less redundancy) in terms of data sharing. The on-going hypothesis is that if we build data-centric systems that rely on ontologies, that can be easily accessed, referenced or manipulated, it can lead us to very clear guidance how to build systems that are not closed and isolated and make their data accessible by a wider range of the applications.

We subscribe to this general hypothesis and are interested in identifying specific advantages of the ontology-based approaches. Namely, by using ontologies, we expect the exposed data, supported by auxiliary tools, can be referenced, read or updated from various systems in a common and unified way. In addition, we expect this would lead to high-quality, and cost-effective data sharing approaches, while addressing many shortcomings of traditional data sharing approaches. We propose an approach that will minimise needs to transform and store

[1] http://www.w3.org/TR/REC-xml/

[2] http://www.json.org/

[3] Comma Separated Value

[4] File Transfer Protocol

data redundantly in local data structures. We propose data-centric systems where data will be published on the internet (we do not consider security issues and other kind of authorisations) and referenced from various systems that use the data, without a need to store them in local data structures.

Research in ontology-based approaches to develop effective tools that will demonstrably cut down the cost of data sharing is continuing. Of special interest are efforts that use linked data networks [6, 7]. In this paper, we review relevant work in the linked data networks research for data sharing, which uses ontologies and provide additional methods and tools to address cost issues in data sharing of enterprise systems.

The paper is structured as follows. First, we present related work focusing on data sharing using ontologies. We continue with a case study that provides a basis for our analysis of requirements for data sharing. Specifically, we analyze the existing research approaches to data sharing, as applied to the case study, and identify their shortcomings as the opportunities for advancement. Then, we describe the proposed linked data network approach to ontology-based data sharing by addressing the case study example problem. Specifically, we analyze the potential of the approach by addressing the identified shortcomings of the existing data sharing approaches. At the end we conclude with a short discussion of the findings and the future work.

II. Related Work

Data sharing in heterogeneous environments continues to be explored by many researchers and various approaches have been analysed and published over the years. As analysed in [12], two main research directions addressing data sharing and data integration are: (1) peer-to-peer approach [1, 8] and (2) global mediation schema approach [2] (often called referent ontology in semantic based approaches [9]). Peer-to-peer approach is relying on many-to-many mappings and transformations where every subject in data sharing process has relation with each other. Global mediation approach is based on a globally defined common schema and a set of rules to transform data structure from a local to a global model. Most of the proposed approaches are trying to solve how to transform data owned by one system to be consumable by another system. We rather want not to store and transform data structure to a local system, but just to reference and use the data straight from their native system.

As observed in [1], an approach with a global mediator schema has a bottleneck mainly because mediation schema must be done carefully and globally while local systems can change significantly, which can violate the mappings to the mediated schema. Another drawback is that changes to global schema may be effected only by a central administrator. Usually, this is a slow and tedious process. This is one of the main reasons why many of the global standards that try to capture all domain variations have failed. To avoid this problem, the authors in [1] propose peer-to-peer architecture called peer data management system (PDMS). The authors emphasise peer-to-peer approach, but not excluding possibility to use global mediation schema as a regular node in peer-to-peer communication, which is also valid for our work. This approach is more focused on writing global queries (global as view) and applying them to local data structure (local as view). Our approach respects peer-to-peer connection too, but differently from [1], we are more focused on not transforming data on the database level, but more on linking data without necessary mappings and transformation.

Frischmuth and collaborators in [2] have investigated complex data integration of integrated systems such as ERP, CRM or SCM. The authors observe that existing SOA architecture may be well-suited for transaction processing, but not for enterprise data integration. They point at Linked Data principles to constitute a promising approach. The authors have identified six crucial areas where data integration challenges arise. One of the areas is Database Integration where the authors also identify that heterogeneous sources require a costly transformation of the data into the relational model. The authors are more focusing their research on federated queries and how Linked Data approach can help to query various data sources using SPARQL [10]. They emphasise that different sources can be integrated in one logical system which is one of the goals of our approach (building common semantic data layer). Also, the authors are not focusing their research on how data from different systems are linked, but more on how to query data from different sources. In [5], authors are investigating ontology usage in data sharing process, but their work is focused on data sharing specifically for social networks. In [11], a data sharing approach in context of Internet of Things is described with architectural solution called Structural Service. This approach is not data-centric, but more service-centric. We want to focus our research on how data should be referenced and shared among different applications.

III. Case Study

In this paper, we describe common problems that exist in a data sharing process between systems which are traditionally treated as closed.

Analysing business applications, we observe that integration problems are tidily connected to data manipulation and transformation from one to another data structure. We can consider a simple example with three parties (e.g, manufacturer (X), distributer (Y), and seller (Z)) sharing data about new product made by the manufacturer. Suppose all of them have closed systems with public API's used for integration. We analyse process in peer-to-peer collaboration manner as well as global mediation schema manner. We do not consider details about type of messages that float among the systems nor the used protocols; we are focused on steps that have to happen in the data sharing process.

A. *Peer-2-peer approach case study*

In Fig. 1, we display typical data flow when three parties collaborate in a data sharing process when peer-to-peer approach is applied.

We simplify necessary communication steps and summarise them in Tab. 1. We summarise just communication between X and Y, but same process repeats between Y and Z and so on.

Let's analyse key points in this approach. We use 'Q' to represent a 'Question' and 'A' to represent a possible 'Answer'.

Step 1: To be able to communicate with Y, system that

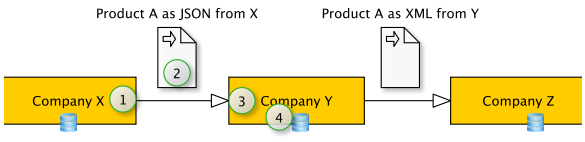


Fig. 1. Peer-to-peer data sharing process

belongs to Company X must implement data structuring

Tab 1. Common steps in data sharing process - peer-to-peer approach

Step	Step description
1	X: prepare data in format that belongs to Y
2	X: send data to Y
3	Y: transform data (use mappings)
4	Y: store data in local database

routines specific to Y (e.g, service calls to Y). It means that X must structure data in a format expected by Y.

Q1: What will happen when X wants to communicate with another company, e.g. W?

A1: X must implement routines to call W API as well. As number of integrated systems grows, number of different implementations grows, too.

Q2: X will send exactly data supported by Y, not worrying about further processing needs. What if data that exists in X are needed by Z, but not communicated between X and Y, because Y does not need them?

A2: Z must implement routines to call X in order to get the needed data.

Step 2: To be able to send data, system X must be aware of connection details of system Y.

Q1: What if connection fails for any reason?

A1: It is system X's responsibility to handle failures even if they belong to other systems.

Q2: What if Y decides to change their API or message format?

A2: X (and all other systems) must re-implement their routines.

Step 3: To be able to use data sent by X, Y must handle data mappings between external system and their internal representation[5].

Q1: What if another system wants to send data to Y?

A1: Y must implement mappings for new system as well.

Q2: What if X decides to change their data or add new ones?

A2: Y must re-implement existing mappings with X or add new mappings for new data.

Step 4: To be able to use data sent by X, Y must store it in a local database after applied transformation.

Q1: Do we have redundancy here?

A1: Yes. Two systems keep same data in two different data structures. As the number of systems grows, redundancy grows, too.

B. Global mediation schema approach case study

Fig.2 represents typical data sharing process when global mediation schema approach is in place. We will not discuss details of this system because there are various approaches how it can be implemented [2, 9]; for our purpose we will just call it Global Mediation System (GMS).

Please note in Fig. 2 steps (3) and (5) are usually done by transformation engine (also called reconciliation engine [13, 14]) which is usually a globally defined tool

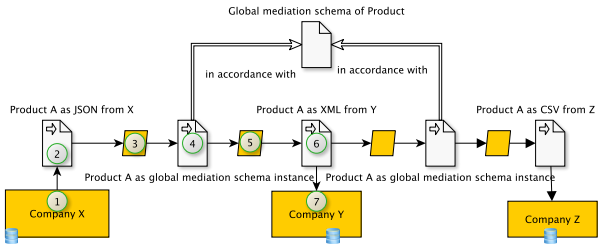


Fig. 2. Global mediation schema sharing process

that simplifies mappings and transformation process. For the simplicity of analysis, Fig. 2 is missing global services which should be used to support communication between GMS and local systems.

In Tab. 2, we list simplified steps that occur in data sharing process when a global mediation schema approach is in place. We suppose all preparation steps are already implemented (like mappings, transformation rules, etc.). Steps 2, 4 and 6, are mostly related to data transfer and protocols and they will be skipped in Tab. 2. System behaviour in these steps is almost same as one we already discussed as Step 2 in Tab.1. Step 7 is same as Step 4 from Tab. 1.

Tab 2. Steps in data sharing process - global mediation schema

Step	Step description
1	X: prepare data in format acceptable by reconciliation engine (3)
3	(3): transform data (use mappings to map codes between X and global mediation schema). Also, use transformation rules to transform from X instance to global mediation schema instance.
5	Reconciliation engine (5) transforms global mediation schema instance to local one

Discussion of steps from Tab. 2 in a question-answer form is giving next.

Step 1: To be able to collaborate using a global mediation schema approach, one has to define local data structure which will be able to map on a global mediation schema.

Q1: What if system X needs to share attribute or structure with system Y which is not supported by global mediation schema?

A1: X and Y have to: (1) wait until global mediation schema supports this and then implement reconciliation on their side, which is time consuming; (2) use flexible fields[6] usually provided by global schema, which introduces complexity in the communication and is out of standard; and (3) implement separate peer-to-peer communication, which is not in sync with GMS nature.

Step 3: To be able to transform data from local data structure to one which is in accordance with global mediation schema, reconciliation engine must take in account mappings related to data structure as well as to data itself.

Q1: What happens when new data appear in local system X, like new product has been inserted in database that belongs to system X?

A1: Additional data mappings must be defined between system X and GMS. Further, additional mappings between GMS and other local systems must occur as well.

[5] Transformation and mappings can be done on side of X as well, but it does not affect our process analyses in a meaningful way.

[6] e.g. http://docs.oracle.com/cd/E28271_01/fusionapps.1111/e15524/flex_gs.htm

Step 5: To be able to transform data from global mediation schema instance to local system Y, reconciliation engine must take in account mappings related to data structure as well as do data itself.

Questions and answers for discussion here are same as ones we have in Step 3.

C. *Requirements for data sharing approach*

As a follow up to the presented discussion related to existing peer-to-peer and global mediation schema approaches for data sharing, we have extracted key requirements that a new approach should ideally satisfy in order to improve data sharing mechanism and techniques. These advancements will lead us towards a common and unified way for data sharing among heterogeneous systems.

Tab. 3 gathers requirements that should be crucial for achieving better data sharing approach, compared to the existing ones. As reference nomenclature, we use T (as reference for tables), S (as reference for steps) and Q (as reference for questions).

IV. Ontology Based Data Sharing Approach

In order to satisfy requirements from Tab. 3, we propose a new approach that uses ontologies to expose data that will be shared. We base our approach on an assumption that once published, data on the network become persistently accessible for various applications at any time, which makes data sharing process easier. From our standpoint, data sharing means publishing data on the network. Differently from traditional approaches that usually rely on public API's or other kind of application layers that hide data, we propose data centric approach, which means direct data-to-data connectivity. Nowadays we have a number of tools[7] and approaches [15] that provide possibility to have data publicly available. Wider list of existing tools can be found in [6]. Usually, places that hold any kind of data are called repositories. If a repository holds ontologies, then it is called ontology repository[8]. In this paper, we will not deeply analyse ontology repositories, but as shown in Fig. 3, ontology repository is an essential part of our approach. The purpose of the ontology repository is to host and manipulate published data. Repository must have at least two major characteristics: (1) ontology storage and (2) management system which will manage hosted ontologies. It is comparable to database approaches, where we have data files and system to manipulate the data (e.g. RDMS). However, in our approach, data files are actually ontologies and management system is administration part of the repository (Repository manager in Fig. 3). Repository manager can have different implementations, but it can be very generic which means a standard set of functions can be supported, like CRUD operations. In our examples below we will assume our Repository manager has public API (e.g. REST Web services[9]) that can be externally called in order to manipulate ontology data (e.g., POST service call http://www.companyX.rs/create will append provided class instance to appropriate ontology). In accordance with Fig. 3, first step in data sharing process is data publishing (Step 1 in Tab. 4). Data publishing is process where source system permanently publishes appropriate data to the ontology repository. We do not consider in details the way how system can provide or transform data to the ontology format, as a number of approaches on this topic have been published [16].

Tab 3. Requirements for ontology data sharing approach

No	Requirement	Source
1	Integrating a new system in data sharing process must be easy and with minimal effort.	T1/S1/Q1, T2/S1/Q1
2	Source system has to be independent of destination system API's or other specifics and vice versa.	T1/S2/Q2
3	Data loss must be avoided.	T1/S1/Q2
4	Once shared data must be permanently available to all subjects in integration process (according to their rights).	T1/S1/Q1, T2/S1/Q1
5	It is source system responsibility to publish data - destination system is responsible to pick up data and handle exceptions if any.	T1/S2/Q1
6	Destination system is responsible to implement all mappings and transformations.	T1/S3/Q1
7	If source system decide to change data structure or data it self, it must not block data sharing process.	T1/S3/Q2, T2/S3/Q1, T2/S5/Q1
8	Data sharing process must generate minimal data redundancy.	T1/S4/Q1
9	If two systems are part of wider integration process, their peer-2-peer data sharing process should not be blocked.	T2/S1/Q1

Research review on this topic can be found in [17]. Published data will get their identifier URI [3], which can be used by any system or application to access or reference this data. If we have permanently published data, we will satisfy requirement No. 4. from Tab. 3. In the text below we will use segments of an example ontology

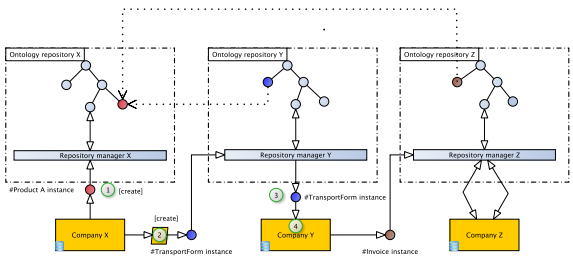


Fig. 3. Ontology based data sharing process

[7] E.g. http://www.stardog.com, www.ontotext.com.

[8] E.g. http://dbpedia.org, http://vocab.deri.ie,

[9] It can be also implemented as messaging system (E.g. JMS)

in order to demonstrate data sharing process displayed in Fig. 3.

```
<!ENTITY example "http://www.companyX.rs/example#" >
…
<owl:NamedIndividual rdf:about="&example;A">
        <rdf:type rdf:resource="&example;Product"/>
        <code>PC-A</code>
        <name>Product Name A</name>
</owl:NamedIndividual>
```

Assuming presented ontology is published (Step 1, Tab 4) and it is hosted and managed by ontology repository available at the address: www.companyX.rs the URI for Product instance named 'A' will be: http://www.companyX.rs/example#A.

Step 2 from Tab. 4 actually means data sharing between Company X and Company Y. Creating data in the Ontology Repositories means data are shared. In our demonstration example, as already explained, Repository manager of Company Y will have a POST service call on URL: http://www.companyY.rs/create that accepts JSON as payload arguments. It means that a system from Company X can create document directly in the Ontology Repository of Company Y. For demonstration purposes we transform OWL ontology to JSON. There exist a number of different tools that can automate this process[10].

```
{"targetOntology": "http://www.compY.rs/example",
"data": {"owl: NamedIndividual": [{
"-rdf: about": "http://www.compY.rs/example#transportForm-123",
"rdf: type": {"-rdf: resource": "http://www.compY.rs/example#TransportForm"}},
{       "-rdf: about": "http://www.compY.rs/example#item-1",
        "rdf: type": {
        "-rdf: resource": "http://www.compY.rs/example#DocItem"},
        "productReference": "http://www.companyX.rs/example#A",
        "quantity": {
        "-rdf: datatype": "http://www.w3.org/2001/XMLSchema#integer", "#text": "1099"}}]}}
```

Ontology fragment displayed above has elements that deserve to be explained:

targetOntology - since service call belongs to Repository Manager which is a generic one and handles different ontologies it is important to explicitly quote which ontology will be affected.

data - data that has to be inserted. In our example it is OWL code fragment converted to JSON. Repository Manager is responsible to convert it back to OWL and append it to the target the ontology.

productReference - demonstrates possibility to have data from one system (#transportForm-123.#item-1) that directly reference data that belongs to another system (#A)[11]. This will minimise data redundancy which will satisfy requirement No. 8 from Tab. 3. Further, if Company X decides to change #Product (which is #A instance of), it will not affect Company Y in the sense of braking their communication channel. Company Y will still be able to accept any instance of #Product since they share only references[12]. This fact is in accordance with requirement No. 7 from Tab. 3.

As demonstrated in the example Company X (source system) is responsible to publish data and Company Y (destination system) is responsible to accept data at any time when it is decided by Company Y. It is in accordance with requirement No. 5. from Tab. 3.With this approach and Ontology Repositories in place source system is not dependent on destination system and their specifics. This is in line with requirement No. 2 from Tab. 3. This requirement is not fully supported since all systems have to agree to use Ontology Repositories as a data sharing layer in their communication.

Tab. 4 Data sharing steps in ontology based approach

Step	Step description
1	X: publish data to ontology repository
2	X prepares and send instance of the document that belongs to Y calling Repository Manager of Y
3	Y has been notified that published ontology is changed by external system (created now instance of #TransporForm document with reference to external product #A)
4	Store data in local database

Step 3 from Tab. 4 emphasises needs for two-fold communication between local system and Repository Manager. Repository Manager must notify local system about any change that happened in ontology. Also, any change that has happened inside the system has to be extended to Repository Manager in order to keep ontologies in sync with the system. It is the local system's responsibility to implement communication between the system and Repository Manager as well as all mappings and transformations that can happen. This is in accordance with requirement No. 6 from Tab. 3.

Step 4 from Tab. 4 is a common step where the local system stores data in a local database. It depends on the local system how it will be done, but we propose that all data that comes from external source should have minimal data redundantly saved in local database. As an example, in Tab. 5, we present how database table (one that keeps data about products) from Company Y should look like.

In short, table should have 'id' as primary key, URI of the record (belongs to ontology where it is published) and other import columns (like 'name'). Table might have fields like 'type' that can mark record as internal (I) or external (E) which can help local system to understand how to handle particular record. As shown in Tab. 5, record that comes from external source should have just minimal data stored (e.g., there is no value for 'name' column). If attribute of the record that comes from external source has to be displayed or used somewhere (e.g. on user interface) it is responsibility of local system to communicate with Repository Manager which will provide needed data using URI.

In our example, if Y needs to share product data to Z it will be again just reference to product that originally comes from system X. There is no possibility to loose data about product side system of Company Z will have direct access to the shared product reference. This will satisfy requirement No. 3 from Tab. 4. Introducing new system to the data sharing process will mean installing another Repository Manager. Of course, new system will need to

Tab. 5. Example database table that collect external and internal data

id	name	URI	type
1	Prod1	http://www.companyY.rs/example#prod1	I
2		http://www.companyX.rs/example#A	E

[10] E.g. http://www.utilities-online.info/xmltojson
[11] On Fig. 3. it is displayed as dotted arrow from Y towards X
[12] It might affect system of the Company Y on some other area (e.g. user interface)

implement communication between Repository Manager and local system. This partially satisfies requirement No. 1 form Tab. 4. Communication between any two systems can be easily extended just introducing new ontologies, which is in line with requirement No. 9 from Tab. 4.

V. CONCLUSION AND FUTURE WORK

We analyzed existing data sharing approaches and grouped them as peer-to-peer and semantic mediation schema approaches. By applying them to a case study, we define a list of requirements that any data sharing approach ideally may support. We proposed linked data network approach to ontology-based data sharing, which relies on a network of ontology repositories. We used the same case study to demonstrate that the proposed approach meets requirements.section to describe our findings.

We will focus our future work to provide detailed definition of ontology repository features as well as to implement Repository Manager that can be used in various data sharing situations.

ACKNOWLEDGEMENTS

We thank Nenad Ivezic from National Institute of Standards and Technology (NIST) for assistance and comments that greatly improved this paper.

REFERENCES

1. A. Halevy, Z. G. Ives, D. Suciu and I. Tatarinov, “Schema Mediation for Large-Scale Semantic Data Sharing”, VLDB Journal : the International Journal on Very Large Data Bases, Volume 14, Issue 1, March 2005, pages 68-83.
2. P. Frischmuth, S. Auer, S. Tramp, J. Unbehauen, K. Holzweißig, and S.-M. Marquardt, “Towards Linked Data based Enterprise Information Integration”, in 'Proceedings of the Workshop on Semantic Web Enterprise Adoption and Best Practice (WASABI), 2013
3. T. Barners-Lee, R. Fielding, L. Masinter, “Uniform Resource Identifier (URI): Generic Syntax”, IETF Network Working Group, 1998
4. H.Wache, et al., “Ontology-based integration of information – A Survey of Existing Approaches”. In Proc. IJCAI Workshop on Ontologies and Information Sharing, pages 108–117, Seattle (WA US), 2001
5. G. Correndo, H. Alani, “Collaborative Ontology Mapping and Data Sharing”, Tech. Rep. No. SO17 1BJ. Southampton, UK: University of Southampton, Electronic and Computer Science Department. Retrieved October 18, 2010
6. C. Bizer, T. Heath, T. Berners-Lee, “Linked Data – The Story So Far”, International Journal on Semantic Web and Information Systems (IJSWIS), Vol. 5, No. 3. March 2009, pp. 1-22
7. T. Heath and C. Bizer: Linked Data, “Evolving the Web into a Global Data Space”, (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool, 2011
8. D. Fensel, S. Staab, R. Studer and F. van Harmelen, "Peer-2-Peer enabled semantic web for knowledge management." Ontology-based Knowledge Management: Exploiting the Semantic Web. Wiley, London, UK, to appear 2002
9. M. Vujasinovic, et. al., “Semantic Mediation for Standard-Based B2B Interoperability”. IEEE Internet Computing”, Vol. 14, No. 1, 2010, pp. 52-63.
10. E. Sirin, Evren, and P. Bijan. "SPARQL-DL: SPARQL Query for OWL-DL." OWLED. Vol. 258. 2007
11. N. Bessis and C. Dobre, “Big Data and Internet of Things: 35 A Roadmap for Smart Environments”, Studies in Computational Intelligence 546, DOI: 10.1007/978-3-319-05029-4_2, © Springer International Publishing Switzerland 2014
12. I. F. Cruz and X. Huiyong, "The role of ontologies in data integration." Engineering intelligent systems for electrical engineering and communications 13.4 2005: 245.
13. N. Preguiça, M. Shapiro and C. Matheson, “Semantics-based reconciliation for collaborative and mobile environments”. In On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE (pp. 38-55). Springer Berlin Heidelberg, 2003
14. ARES - Athena Interoperability Framework. On-line: http://athena.modelbased.net/solutions/singular_solutions/solution_ares.html
15. E. Hyvönen, K. Viljanen, J. Tuominen and K. Seppälä, “Building a national semantic web ontology and ontology service infrastructure–the FinnONTO approach”, (pp. 95-109). Springer Berlin Heidelberg, 2008
16. N. Konstantinou, D-E. Spanos and N. Mitrou. “Ontology and database mapping: a survey of current implementations and future directions”, Journal of Web Engineering, 7(1), 001-024, 2008
17. V. Jain, “Mapping Between RDBMS And Ontology: A Review”, International Journal of Scientific and Technology Research, Vol 3, Issue 11, November 2014

Simulation of tariff plan selection by online users using Agent Based Models

Aneesh Zutshi*, Tahereh Nodehi**, Ricardo Jardim-Gonçalves** and Antonio Grilo*
* DEMI, FCT, Universidade Nova de Lisboa,
** DEE, FCT, Universidade Nova de Lisboa,
aneesh84@campus.fct.unl.pt, t.nodehi@campus.fct.unl.pt, rg@uninova.pt, acbg@fct.unl.pt

***Abstract*— Online Businesses can be represented as a complex interaction of interconnected online users responding to the value proposition of an online company. We propose an Agent Based Modeling framework (DYNAMOD) that aims to explain these complex dynamics. This framework aids in the creation of simulation models that mimic the actual market behavior and perform business forecasting and decision support functions. Through a case study of the largest e-procurement provider in Portugal – Vortal.biz, we simulate their pricing model and its impact on user behavior and revenue.**

I. INTRODUCTION

Understanding the new digital economy has been a challenge for companies that were tuned to the traditional ways of doing business and were armed with traditional product development and marketing philosophies. Today Digital Business managers often rely on experience and intuition to set up business models and pricing strategies. Though marketing surveys, have been traditionally used in gauging consumer willingness to pay, a big challenge is to predict business growth, customer adoption and customer response to specific business and pricing models [1], [2]. Online Businesses are complex interacting systems where online users interact and share opinions and experiences at a rate far greater than traditional brick and mortar businesses. These user opinions are shared through offline and online word of mouth (WOM) channels, in addition to various marketing channels. Customer Satisfaction is a key to spread of positive or negative WOM, which influences new customer adoption rate. System Dynamics have often been used to model consumer adoption [3]. However System Dynamics require the rules of the behavior to be written at a higher level, such as how the whole population of consumers will respond to a marketing activity rather than how a particular individual will respond [4].

Online businesses are examples of a complex system, where the behavior of individual users can be used to model the growth or decline of a business proposition [5]. This could provide an analytical approach to develop models that can be used by businesses as a decision support system. Such models can perform a range of objectives, such as making business forecasts, calculating the implications of a change in product pricing, optimization of different price plans, and simulating the impact of a change in the Business Model.

Agent Based Models have been chosen as the most appropriate tool to implement this complex systems approach. Agent Based Modeling is a new computational method through which macro-level consequences are explained through simplified representation of micro level interactions between agents that represent real life entities [6]. These autonomous agents represent online users with individual characteristics as well as independent internal decision making capabilities.

In this paper, we propose a Dynamic Agent Based Modeling Framework (DYNAMOD), that incorporates Agent Based Modeling (ABM) techniques to develop and test digital business models in a variety of online market scenarios. Further, this paper explores its applicability to forecast and simulate changes in the Pricing Models for the largest Portuguese e-procurement platform provider – Vortal.biz.

II. THE MODELING SCENARIO

Vortal.biz is the largest procurement platform provider in Portugal and has a market penetration close to 90% of its assessed potential market. Such a high penetration was possible due to the use of a freemium pricing model and the ability to attract a high percentage of large buyers. Vortal´s platform is an example of a double sided platform that tries to attract large volume buyers on one side and sell monthly subscription plans to various vendors on the other side. The buyers include both public and private entities. Contracts with buyers are often negotiated on a case to case basis with large discounts being offered to bulk buyers to make the platform more attractive for sellers. Sellers are offered the choice of choosing between a range of tariff plans. This paper only focuses on creation of an Agent Based Simulation Model for the sellers. This is firstly because contracts with buyers are more arbitrary in nature and hence difficult to model and secondly, in a short time period of one or two years, the number of buyers do not significantly change due to a saturated market penetration by Vortal. Hence the business may be modeled as a single sided platform despite being a double sided platform in reality.

A. Tariff Plans

The Sellers are offered a freemium tariff plan with the Universal plan as a basic free plan. Smart Plans offer 4 advanced features while Best Plans offer 4 more

advanced features. Universal (U), Smart-Gov (SG) and Best-Gov (BG) plans offer clients access to opportunities only from the Public entities while Smart-Eco (SE) and Best-Eco (BE) also offer opportunities for the Private as well as public entities. The Sellers are charged a Monthly Tariff based on the company size according to which four Tiers of Tariff are offered. However since the largest number of clients are in Tier 1, we will consider only Tier 1 in this simulation. Similar simulation models can be prepared for other Tiers.

B. Model Objectives

The model seeks to simulate the customer response to different plans and offerings in order to develop an agent based simulation model that can simulate the impact of changes in pricing of the various Tariff Plans. Also, the company was interested in viewing the attractiveness of offering users the option of using one of the premium features without upgrading to the higher tariff plan. Such a scenario was also modeled in this simulation. To summarise, the specific objectives of the model are as follows:

1. Simulate the upgrade of customers to newer tariff plans.
2. Simulate the subscription of a single feature instead of upgrading to a newer tariff plan.

III. THEORETICAL BACKGROUND

The development of the DYNAMOD model is based upon other previous research works in diverse areas. Some key concepts that have been applied in the model are discussed below

A. Use of Agent Based Models for Solving Business Problems

New tools and techniques are necessary to help model the complex nature of online products and services. Hence we need to develop a customizable simulation environment that can capture the dynamics of an online market, and provide Business Managers with tools to simulate and forecast, thus aiding to perfect their Business Model. Online markets can be represented as a network of interconnected online users which share positive and negative feedbacks and respond to different online products and services. If the behavior of individual agents can be sufficiently well modeled, then a natural candidate for representation is multi-Agent Based Modeling Techniques.

ABM is build on proven, very successful techniques such as discrete event simulation and object oriented programming [7]. Discrete-event simulation provides a mechanism for coordinating the interactions of individual components or “agents” within a simulation. Object-oriented programming provides well-tested frameworks for organizing agents based on their behaviours. Simulation enables converting detailed process experience into knowledge about complete systems. ABM enables agents who represent actors, or objects, or processes in a system to behave based on the rules of interaction with the modelled system as defined based on detailed process experience. Advances in computer technology and modelling techniques make simulation of millions of such agents possible, which can be analysed to make analytical conclusions [8].

The literature review reveals that applications of ABM have been made to model specific areas of Business. These include prediction of financial distress [9], product adoption [(S. Kim et al. 2011), [11], [12], [13]], consumer behaviour [[14], [15]], market share [16], Urban Management [17] and demand forecasting [18]. [16] demonstrated the possibilities of predicting market share based on certain BM attributes of Frontier Airlines. [19] addresses the issue of capturing Internet behaviour to deliver relevant advertisements. ABM approaches can also be used for modeling user response to different sources of advertising. It can also be used to model response to identify the most critical target groups, complementing traditional approaches for the same [20].

[21] propose an Agent Based Model to simulate consumer decision making based on culture, personality and human needs and relates them to car purchase decisions. Tesfatsion introduced Agent-Based Computational Economics (ACE) as the computational study of dynamic economic systems modeled as virtual worlds of interacting agents. [22] have applied ACE to retail and wholesale energy tradings in the Power Markets. In this paper we extend the concept of Agent-Based Computational Economics, to develop DYNAMOD- An Agent Based Modeling Framework for online Digital Business Models.

IV. THE VORTAL ABM MODEL

A. The generic DYNAMOD Model

The Vortal Model is based on the customization of the DYNAMOD Simulation model that has been applied to a variety of online business case studies before. The DYNAMOD Framework has been developed based on the academic literature collected regarding the unique aspects of an online business. Its purpose is to provide researchers and companies engaged in online businesses with a tool for quickly developing Computational Modeling Systems that can represent their Business Models and their Business Environment, in order to perform advanced simulations for predicting business growth dynamics. DYNAMOD is based on Agent Based Modeling, which enables dynamic representation of the online marketplace. Every Buyer that is a customer for a product or service is represented as an Agent in DYNAMOD. These agents interact with each other and share information about new products and services. At the same time, they are influenced by Advertising and Social sites. The model captures these influences, and simulates their impacts in order to predict future scenarios.

The model is customizable and extendible to implement a diverse set of Business Model components, and to make a variety of simulations. The model core consists of many interacting agents that represent a market. The model includes standard variables and logics

for implementing influence and satisfaction scores for each agent. This core component handles the simulation and interaction, and defines what constants are needed to initialize the key features of the model.

Other features are added to the model in the form of modules, as and when necessary, for different case scenarios. In the current scope of the model, four additional modules have been envisaged, namely Competitor Analysis, Pricing Analysis, Network Effects/Viral Marketing Effects, and Market Segmentation/Region Based Modeling.

Diffusion of Innovation literature has used two major forms of adoption functions [23]. In the Bass like model, adoption occurs through individual innovation or through peer imitation[24]. In the threshold model, each user adopts only when a certain threshold of its neighbors have adopted[25].

B. *Formalization of the Simulation Model*

Client Upgrades

The model tries to simulate the upgrade of users to higher tariff plan based on their willingness to pay. The upgrade choice for users is based on the assumption that a user will only choose to upgrade one level up at a time, and a user on a plan with access to public as well as private opportunities will not like to move to a plan that restricts opportunities for him to only the public opportunities. Hence, the Universal users may upgrade to Smart-Gov or Smart-Eco, Smart-Gov users may choose to move to Smart-Eco to expand their markets or may choose to move to Best-Gov plan if they would like to continue access to the same public opportunities but with enhanced features. Similarly users may upgrade from Smart-Eco as well as Best-Gov to Best-Eco. (See Figure 1)

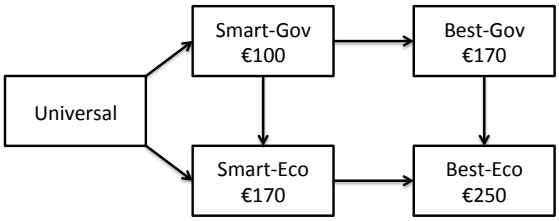


Figure 1. Client Upgrade Options

To determine if a user is willing to upgrade to a higher plan or not, we need to identify the user willingness to pay for the higher plan. This was determined through sample online surveys where each user was asked about his interest and willingness to pay for the tariff plan options that he is most likely to upgrade to.

For each User Agent A_i,

The current plan, A_i [plan] = 0, 1, 2, 3, 4, 5 corresponding to Universal, Smart-gov, Smart-eco, Best-gov, Best-eco.

Willingness to Pay for an agent A_i currently on Plan n for Plan m = A_i [WTP – n – m]

Current Price for Plan n = Price-n

An agent with A_i on universal plan will upgrade to Smart-Gov if it is willing to pay more than the current price.

If (A_i [WTP-0-1] >= Price-n) , A_i [Plan] = [0 ➔ 1]

In case an agent is eligible to upgrade to more than one higher plan, he will upgrade to the plan that he sees a higher value, ie, his willingness to pay is higher.

For any optimized pricing model, the company is interested in maximizing the Total Monthly Revenues.

Total monthly revenue = R = nSG * Price-1 + nSE * Price-2 + nBG * Price-3 + nBE * Price-4

Where nSG, nSE, nBG and nBE are the number of Agents currently on plan Smart Gov, Smart Eco, Best Gov and Best Eco respectively.

C. *Optional Features*

Vortal was interested in using this model to simulate the impact of additional revenue generation models. One of the possibilities included the introduction of users to chose one optional premium feature from a higher plan without needing to upgrade to a higher plan. The features are listed as follows:

{f1, f2, f3, f4} are offered in plans SG, SE

{f5, f6, f7, f8} are offered in plans BG, BE

A user at plan U does not have access to any of the features. He may chose to subscribe to f1 or f2 or f3 or f4 but not to more than one of them at the same time. Similarly one of f5, f6, f7 or f8 is available to a user of SG or SE Tariff plan. Features must be priced in a way such that they do not cannibalise the upgrades to a higher tariff plan. The set of possible monthly prices for the eight features are:

{f1, f2, f3, f4} can have the following prices {€40, €60, €80, €100}

{f5, f6, f7, f8} can have the following prices {€30, €40, €50, €60}

An agent with A_i on universal plan will upgrade to f1 if it is willing to pay more than the current price of f1. If this is true the flag f1 for the agent A_i is set to 1.

If (A_i [WTP-0-f1] >= Price-f1) , A_i [f1] = [0 ➔ 1]

The users willingness to pay for different features were again collected through the sample survey. User's willingness to pay for each of the features were evaluated and a distribution of the same was used for programming the agents in the simulation model. If an agent's willingness to pay is higher for more than one feature, and he is not ready to make the switch to a higher tariff plan, then the feature for which his willingness to pay is the highest shall be adopted to. If he has the same willingness to pay for more than one feature, then the order of utility for features according to the overall utility of various features based on the sample survey.

D. *Data Collection for the Vortal Model*

An Agent Based Model mimics the real life consumer opinions and preferences and uses these to simulate the larger business scenario. To gauge user opinions, preferences and willingness to pay for various tariff plans, we conducted an online questionnaire to gauge the user satisfaction, influence from various sources of advertising and willingness to pay for upgrade to a higher plan. The questionnaire was sent to 7000 randomly selected clients. The total number of respondents were 365 (U – 290, SG- 37, SE -19, BG- 15, BE-4) which is a

response rate of more than 5%. The distribution of respondents according to the maximum willingness to pay for an upgrade is shown in

Similarly, the willingness to pay for each of the 8 features was obtained from the respondents and arranged according to the current tariff plan that they are currently on. These were fed into the Simulation Model for the purpose of simulating user adoption of feature based pricing.

E. NETLOGO – The Agent Based Modeling Tool used

The DYNAMOD model is generic and can be modeled using any modeling language. However we have used NETLOGO 5.0.3 as the development environment to model this case study. This provides us with an extensive library to program the agent behavior, environmental constraints and the modeling parameters. It also provides us with a graphical interface to review the simulation results.

NetLogo is a freely download- able, agent-based software package that was created at the Center for Connected Learning and Computer-Based Modeling (CCL), directed by Uri Wilensky, at Northwestern University. NetLogo utilizes a simple programming language and a convenient user interface allowing models to be easily simulated and evaluated. The software application is designed to be easy to use yet is broadly utilized by academics in the social, computer, and "hard" sciences. NetLogo is extremely flexible. Interactions not only among autonomous agents, but also between agents and the simulated environment, can be specified.

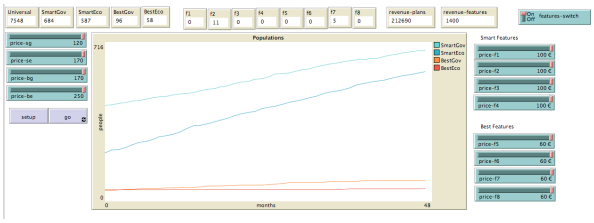


Figure 2. Netlogo ABM Model Interface

DYNAMOD Models are based on output of time series data, so that customer adoption and preferences are modeled over a period of time. The concept of time is modeled in NetLogo as an arbitrary unit (referred to as a "tick") that is left to the model developer to define. In our model, a tick is recorded after every consumer agent makes a random "walk" within the simulated environment. These consumer movements are what enable interactions with other consumers or brands. The total number of time units per iteration, as well as how many units represent a single year, can be modified by the user to best represent a specific study context. Figure 2 shows the Netlogo Model for this case.

V. MODEL VALIDATION

To validate the model we obtained the actual number of users for each price plan in the end of 2012 and 2013 from Vortal (TABLE I). We initialized the model to have different agent types based on the number of users on different tariff plans in 2012. The user willingness to pay were attributed to the agents based on the survey results.

TABLE I

	U	SG	SE	BG	BE	Total
2012	8227 91.68%	431 4.80%	218 2.42%	46 0.51%	51 0.56%	8973
2013	7888 84.75%	679 7.29%	458 4.92%	167 1.79%	115 1.23%	9307

The model is allowed to run till it reaches the closest to the real value of number of SG users in 2013. Hence the real value of SG in 2013 is used to determine the average number of steps that constitute a one year period for the simulation model. The model was run 1000 times to monitor the stochastic variability of the results. This enables us to visualize a probability distribution of the forecasts. The average number of steps the ABM took to just cross the number 679 of SG users in 2013 was 60. The 60 steps represented a time scale of one year and the model was run 1000 times for 60 steps and other parameters were measured. Hence, we shall use the forecasts for agents on other tariff plans to validate the accuracy of forecasts of the model for 2013 (Figure 3).

The price of the optional features were set at the maximum possible value, ie, €100 and €60 for the first four and last four features respectively. The simulation results showed that in one year, an average of 11.48 users and 8.96 users subscribed the f2 and f7 features respectively. There were no subscribers for any of the other features. Thus this price point was set too high. We shall discuss further how an optimization of prices can be achieved using genetic algorithms.

The model represents a highly saturated market condition, where there is a high market penetration. Already all major vendors in Portugal are using the Vortal Platform. The model forecasts a strong shift from the free users of Vortal towards becoming paying customers as well as customers upgrading from the "Smart" Plans to the "Best" Plans. It further closely forecasts the percentage of users in various plans with an error of less than 1% in all the cases. However this error is negative across all the paid plans thus showing a small bias where the model predicts a lower willingness to pay than in reality. This bias may be corrected in future research work by implementing other means of gauging willingness to pay other than direct questionaires, for example, Vickery Auction [26]. The prediction capabilities of this model cannot be compared with any other existing forecasting tool, because no forecasting tool enables future prediction when only one set of data for the current period exists [27]. The DYNAMOD model is not just a forecasting tool, but rather a toolkit that enables us to capture the underlying business scenario in the form of user satisfaction and willingness to pay end enables us to make future forecasts based on a variety of what-if scenarios.

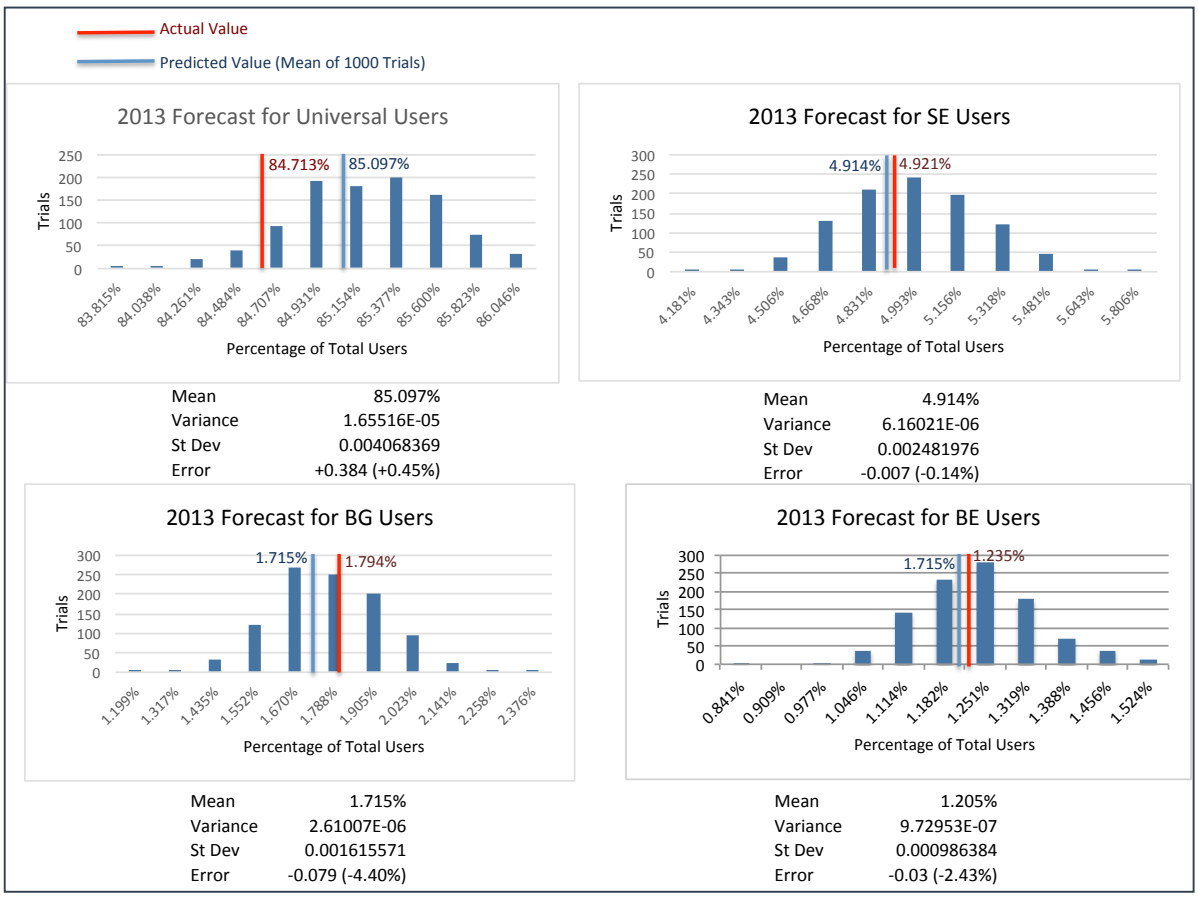


Figure 3. Comparison of actual and predicted value for 2013

VI. Conclusion and Future Work

The objective of the DYNAMOD Modeling Framework is to model the complex dynamics of online users and how they respond to the value proposition of a Digital Business. Through a complex system approach we propose an Agent Based Modeling Framework that can be used as a basis to develop customized Agent Based Models for different online business scenarios. This framework models the most common features of online markets, such as Word of Mouth, Network Effects, Price Sensitivities, Viral Marketing and various sources of User Influences. In this paper we apply this model to develop pricing models for a B2B procurement platform provider.

We demonstrate how user willingness to pay can be measured and used to initialize an ABM model that can then be used to forecast how users will react to different combinations of pricing for different plans and features. We also demonstrate the use of this tool to test the revenue implications for new business scenarios, such as the introduction of new feature based pricing options. This approach provides us with a useful tool to test and optimize business and pricing decisions before they are finally implemented.

The DYNAMOD framework captures the key characteristics of online businesses, however each business is unique and hence this modeling framework must be customized to include specific Business Requirements and Modeling Objectives. The framework also provides a structured approach to developing questionnaires to gauge sample user behavior. Since user behavior changes over a period of time, we are in the process of developing a Decision Support Toolkit that can conduct automated periodic surveys to update user responses and thus adjust the behavior of the agents used in the model to improve accuracy of long time intervals, and alert management about changes in key metrics like customer satisfaction, word of mouth, in addition to simulating its impact.

The complex systems approach is novel in terms of viewing the Business Model of a company and the online networks of users as one complex system, where overall behavior can be understood by capturing individual level behavior. We believe this approach could open up new opportunities for digital entrepreneurs and business managers, and deepen academic understanding about how online businesses work.

References

[1] A. Zutshi, A. Grilo, and R. Jardim-Goncalves, "The business interoperability quotient measurement model," *Comput. Ind.*, vol. 63, pp. 389–404, 2012.

[2] A. Grilo, A. Zutshi, R. Jardim-goncalves, and A. Steiger-garcao, "Construction collaborative networks: the case study of a building information modelling-based office building project," *Int. J. Comput. Integr. Manuf.*, vol. 26, pp. 152–165, 2013.

[3] J. Stermann, "Business Dynamics–Systems Thinking and Modeling for a Complex World," *Irwin McGraw—Hill, Bost.*, 2000.

[4] W. Rand and R. T. Rust, "Agent-based modeling in marketing: Guidelines for rigor," *Int. J. Res. Mark.*, vol. 28, no. 3, pp. 181–193, Sep. 2011.

[5] A. Zutshi, A. Grilo, and R. Jardim-Gonçalves, "A Dynamic Agent-Based Modeling Framework for Digital Business Models: Applications to Facebook and a Popular Portuguese Online Classifieds Website," *Digit. Enterp. Des. ...*, pp. 105–117, 2014.

[6] A. Zutshi, A. Grilo, and R. Jardim-Goncalves, "An Agent Based Modelling approach to Digital Business Models: through a literature review of three complementary Research Areas," in *5th KES International Conference on Intelligent Decision Technologies*, 2013, p. 48.

[7] M. North and C. Macal, *Managing business complexity*. 2007.

[8] A. Zutshi, A. Grilo, and R. Jardim-Gonçalves, "DYNAMOD—An Agent Based Modeling Framework: Applications to Online Social Networks," *Proc. Eighth Int. Conf. Manag. Sci. Eng. Manag. Adv. Intell. Syst. Comput.*, vol. 280, 2014.

[9] Y. Cao and X. Chen, "An agent-based simulation model of enterprises financial distress for the enterprise of different life cycle stage," *Simul. Model. Pract. Theory*, vol. 20, no. 1, pp. 70–88, Jan. 2012.

[10] S. Kim, K. Lee, J. K. Cho, and C. O. Kim, "Agent-based diffusion model for an automobile market with fuzzy TOPSIS-based product adoption process," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7270–7276, Jun. 2011.

[11] J. Diao, K. Zhu, and Y. Gao, "Agent-based Simulation of Durables Dynamic Pricing," *Syst. Eng. Procedia*, vol. 2, pp. 205–212, Jan. 2011.

[12] M. E. Schramm, K. J. Trainor, M. Shanker, and M. Y. Hu, "An agent-based diffusion model with consumer and brand agents," *Decis. Support Syst.*, vol. 50, no. 1, pp. 234–242, Dec. 2010.

[13] S. a. Delre, W. Jager, T. H. a. Bijmolt, and M. a. Janssen, "Targeting and timing promotional activities: An agent-based model for the takeoff of new products," *J. Bus. Res.*, vol. 60, no. 8, pp. 826–835, Aug. 2007.

[14] L. Vanhaverbeke and C. Macharis, "An agent-based model of consumer mobility in a retail environment," *Procedia - Soc. Behav. Sci.*, vol. 20, pp. 186–196, Jan. 2011.

[15] T. Zhang and D. Zhang, "Agent-based simulation of consumer purchase decision-making and the decoy effect," *J. Bus. Res.*, vol. 60, no. 8, pp. 912–922, Aug. 2007.

[16] J. R. Kuhn, J. F. Courtney, B. Morris, and E. R. Tatara, "Agent-based analysis and simulation of the consumer airline market share for Frontier Airlines," *Knowledge-Based Syst.*, vol. 23, no. 8, pp. 875–882, Dec. 2010.

[17] L. Gao, B. Durnota, Y. Ding, and H. Dai, "An agent-based simulation system for evaluating gridding urban management strategies," *Knowledge-Based Syst.*, vol. 26, pp. 174–184, 2012.

[18] Y. Ikeda, O. Kubo, and Y. Kobayashi, "Forecast of business performance using an agent-based model and its application to a decision tree Monte Carlo business valuation," *Phys. A Stat. Mech. its ...*, vol. 344, no. 1–2, pp. 87–94, Dec. 2004.

[19] S. Bellman, J. Murphy, S. Treleaven-Hassard, J. O'Farrell, L. Qiu, and D. Varan, "Using Internet Behavior to Deliver Relevant Television Commercials," *J. Interact. Mark.*, vol. 27, no. 2, pp. 140–130, Feb. 2013.

[20] H. Lee, H. Shin, S. Hwang, S. Cho, and D. MacLachlan, "Semi-Supervised Response Modeling," *J. Interact. Mark.*, vol. 24, no. 1, pp. 42–54, Feb. 2010.

[21] O. Roozmand, N. Ghasem-Aghaee, G. J. Hofstede, M. A. Nematbakhsh, A. Baraani, and T. Verwaart, "Agent-based modeling of consumer decision making process based on power distance and personality," *Knowledge-Based Syst.*, vol. 24, no. 7, pp. 1075–1095, Oct. 2011.

[22] A. Somani and L. Tesfatsion, "An agent-based test bed study of wholesale power market performance measures," *Comput. Intell. Mag. IEEE*, vol. 3, no. 4, pp. 56–72, 2008.

[23] F. Stonedahl and W. Rand, "Evolving viral marketing strategies," *12th Annu. Conf.*, 2010.

[24] F. M. Bass, "A New Product Growth for Model Consumer Durables," *Manage. Sci.*, vol. 15, no. 5, pp. 215–227, 1969.

[25] H. Gatignon, "A propositional inventory for new diffusion research," *J. Consum. Res.*, 1985.

[26] K. Blumenschein and M. Johannesson, "Hypothetical versus real payments in Vickrey auctions," *Econ. Lett.*, 1997.

[27] B. Rossi and T. Sekhposyan, "Understanding models' forecasting performance," *J. Econom.*, vol. 164, no. 1, pp. 158–172, 2011.

IoT Lab Crowdsourced Experimental Platform Architecture

Stevan Jokic*, Aleksandra Rankov*, Joao Fernandes**, Michele Nati***, Sebastien Ziegler°, Theofanis Raptis°°, Constantinos M. Angelopoulos°°°, Sotiris Nikoletseas°°, Orestis Evangelatos°°°, Jose Rolim°°°, Srdjan Krčo*

* DunavNET, Novi Sad, Serbia
** Alexandra Instituttet A/S, Aarhus, Denmark
*** Institute for Communication Systems, University of Surrey, Guildford, UK
°Mandat International, Geneva, Switzerland
°°Computer Technology Institute & Press Diophantus, Patras,Greece
°°° Université de Genève, Geneva, Switzerland
aleksandra.rankov@dunavnet.eu iotlab@genevaproxy.com

***Abstract* — This paper presents the architecture of the crowdsourced experimental platform called IoT Lab. The platform will provide a new approach for experimentation by extending existing IoT FIRE testbeds, traditionally built from static sensor mote platforms, with crowd sourced resources and thus will enable richer and more distributed multidisciplinary experiments with more end-user interactions, flexibility, scalability, cost efficiency and societal added value.**

I. INTRODUCTION

Exploring direct interactions with the crowd through crowdsourcing and crowdsensing techniques while enabling the crowd to be at the core of the research cycle with an active role in research, from its inception to the results' evaluation, is the main motivation behind the development of the IoT Lab platform - testbed as a service (Figure 1) developed by IoT Lab European research project (www.iotlab.eu). This will provide a better alignment of the research with the society, end-users needs and requirements.

Crowdsourcing is recognised as a practice of obtaining needed services, ideas, or a content by soliciting contributions from a large group of people and especially from an online community, rather than from traditional employees or suppliers.

The crowdsourcing approach can apply to a wide range of activities including crowdsourcing of data, measurements, opinions, solutions and funding. The main focus here is on IoT-related, attractive types of crowdsourcing such as collection of data/measurements and user rates/opinions. IoT today regards every smartphone user as a source of data that is generated and shared via his/her device.

There are already a number of crowdsourcing platforms focusing on different ways to empower user participation in the IoT through their mobile phones. While some of the existing platforms might be too specific [1] and only support pre-defined tasks without a possibility to extend them due to the lack of open source code, others might be too general, like Ushahidi [2] that does not support the involvement of participants and only leverages on geo-localized data collection and visualization through maps.

Figure 1 IoT Lab 'Testbed as a Service'

Platforms like Phonelab [3], provide a model for crowd engagement in the IoT co-creation effort, but no actual application is provided to support this, while on the other hand custom applications are developed and distributed according to the selected use case.

There is a number of existing platforms that properly support user participations in IoT experimentation through mobile phones [4][5] but still lack the ability to fully support the IoT Lab envisioned models for participation and interaction between participants and investigators.

Supporting scripting and crowdsourcing on mobile phones is possible through APISENSE [6] but an integration with other IoT Lab provided tools is needed, such as Resource Management and Experiment Management. However, an official version of the platform has not been released yet.

Similarly, a set of other existing platforms, fulfilling different needs envisioned by the IoT Lab mobile app, will be investigated further in order to understand how and to what level they could be extended with other IoT Lab tools, so as to achieve a complete IoT Lab platform integration.

EpiCollect [7] can be useful for creating a survey and questionnaire, but the lack of open source code limits the possibility of extension and integration.

mCrowd [8] seems to better fit participatory sensing applications. However, the lack of APIs and only provided an iPhone version might limit the possibility of integration and extension.

Funf [9] and AmbientDynamix [10] represent good frameworks for crowdsensing and crowdsourcing. In particular, the capability of AmbientDynamix to adapt

its behaviour to context could allow support of different experiment participation models suitable for the end user perspective. The possibility to integrate it with other envisioned IoT Lab resources should be further investigated.

Together with AmbientDynamix, the scope of results is limited to the IoT Lab mobile application. McSense [11] seems to support all the basic functionalities envisioned in the IoT Lab mobile app (actuation and sensing), of which the IoT Lab platform should be comprised, and also includes tools for resource selection, task (and experiments) description and other useful functionalities. The possibility to integrate and extend it with other IoT Lab resources, such as integration with FIRE testbeds Profile Management and the Search and Communication tools, should be investigated further.

For all these reasons, the platforms that should be selected and further analysed to fully understand their potential integration with in the final IoT Lab platform are represented by AmbientDynamix and McSense.

None of the platforms actually foresees the possibility to integrate smartphones with existing FIRE testbeds or in general statically deployed IoT resources, such as smart power meters, home automation control systems and so on. Nonetheless, no effort has been put towards providing virtualization tools that enable heterogeneous IoT resources to be homogeneously available and interoperable with each other. To this purpose, the approach proposed by the IoT Lab platform is new and advanced with respect to existing crowdsourcing platforms.

The paper is organized as follows: Section II presents the IoT Lab vision and discusses the main challenges; Section III describes the architecture design approach whilst Section IV gives details about the architectural components. The sequences of using the services are given in Section V. The paper concludes in Section VI.

II. IoT Lab vision and key challenges

IoT Lab platform aims to enhance the existing IoT FIRE testbeds, traditionally comprising of static sensor mote platforms, by utilizing end-user participants' smartphones and relevant mobile/portable devices in order to achieve crowd participation in sensing and actuation operations and thus enable a wide range of multidisciplinary experiments.

To achieve this aim, the IoT Lab platform addresses the following objectives:

- Crowdsourcing & crowdsensing mechanisms and tools;
- Crowd-driven research;
- Virtualization of crowdsourcing and testbed components;
- Ubiquitous interconnection and cloudification of the testbeds' resources;
- Testbed as a Service;
- Multidisciplinary experiments;
- End-user and societal value creation;

Through these objectives the IoT Lab platform can achieve its goals in terms of (i) connecting and using existing IoT testbeds which increases the testbed economic sustainability; (ii) proactively involving participation of the public through crowdsourcing, as well as researchers taking part in the IoT experiments which provides closer interactions between experiments and society.

The platform also considers issues such as privacy and personal data protection through a 'Privacy by Design' approach and built-in anonymity.

There are various stakeholders identified for the IoT Lab "Testbed as a Service" platform with the main ones being: researchers/experimenters; testbed owners; crowd/media; EC; official authorities and potential private customers.

The expected IoT Lab impact is to support experimentally driven research, in particular to conduct multidisciplinary investigation of key techno-social-economic issues (i.e. Internet Science), to further exploit any relevant FIRE facilities, to consider benefits for citizens as well as to investigate ethical and self-sustainability aspects of experimental facilities.

The key challenge of the IoT Lab platform is to successfully attract researchers and the general public (crowd) into using the testbed facilities and joining the experiments respectively. A range of incentive and rewarding schemes has also been considered. Furthermore, a wider audience needs to be reached throughout the duration of the project both for IoT Lab sustainability reasons and in order to help the platform mature.

Additional technical challenges originate from the IoT Lab platform development process that needs to address heterogeneous identified requirements and to support challenging use cases proposed by the crowd.

III. Architecture design approach

The preliminary IoT Lab platform architecture design is addressing double challenge: On one hand, it has to integrate diverse IoT-related testbeds located in different regions of Europe. On the other hand, it has to integrate smart phones with existing FIRE testbed infrastructures, thus representing a novel approach with respect to existing crowdsourcing solutions. The key platform components have been identified and their functionalities, interaction patterns, interfaces and communication links described.

The derivation process followed an IoT-A methodology [12] to support interoperability and scalability and to enable use of a wide range of heterogeneous devices and testbeds from different application domains thus satisfying a high number of requirements.

An architecture generation process starts with the analysis of technical and end user related requirements derived from selected use cases. Two scenarios have been proposed. The first one is a "Game and Supermarket Marketing" - a smart city scenario in

which users can play a game and participate in experiments, for instance a market survey from a supermarket. The second scenario is "Energy Efficiency and User Comfort Hints" in which crowdsensing methods are used to adapt energy efficiency to human presence and behavior. Moreover, the users can provide feedback on the current devices' setup and set their user preferences representing a Physical Testbed Scenario. These two scenarios have been analysed using the IoT-A methodology and they are here represented using a general use case diagram shown in Figure 2.

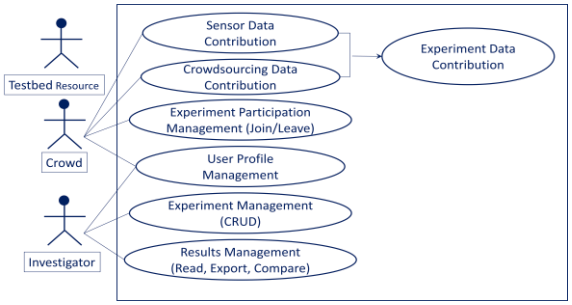


Figure 2 IoT Lab General Use Case Diagram

The analysis of the use cases provided a detailed list of requirements for the IoT Lab platform that can be summarised as follows:

- User profile management for both participants and investigators.
- Experiment Configuration – Investigators must specify a detailed description of their experiments, including needed resources, ethics and privacy concerns, timeline and overall objectives of the experiment.
- Testbed Resources Management – A simple and easy interface that allows testbed managers to configure and make their resources available for experimentation.
- Crowdsourcing and Crowdsensing provisioning – The platform must provide ways that allow both crowd and testbed resources to provide data which can be sensory data and/or crowd knowledge.
- Experiments Management – The system must provide ways that allow the users to browse and evaluate existing experiments and in the case of the investigators to manage their own experiments.
- Privacy and Ethics – Privacy by design concept is followed; users are requested minimal information and for each of the experiments a clear description of the required data (user and device) is presented. Experiments must be validated before being run.
- Support for incentives – Incentives scheme for crowd and investigators need to be included.

Development of the polyvalent and flexible IoT Lab platform 'Testbed as a Service' that can support a large set of IoT related experiments is guided by several considerations:

- Adopting a modular architecture, that will enable the evolution of individual components without impacting the whole architecture;
- Favouring generic enablers that can be easily used by different experiments;
- Aligning with main stream standards and solutions to ease the integration with third parties resources;
- Satisfying the requirements derived from the most up-voted use case scenarios proposed by the consortium and described above.

Selected real use cases for implementation identified several important experimental approaches including crowd sensing; data collection & processing from different IoT testbeds; the code execution on the participant side and completion of different questionnaires by participants.

The generic IoT Lab 'Testbed as a Service' enablers are illustrated in Figure 3.

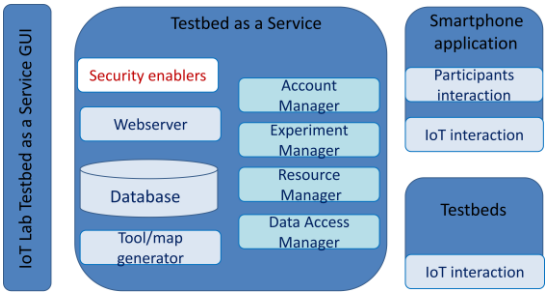


Figure 3 IoT Lab enablers – generic

IV. ARCHITECTURAL COMPONENTS

Main architectural components of the proposed IoTLab system are organised in four groups as shown in Figure 4: Account manager, Resource manager; Experiment manager and Web user interfaces.

A. Account Manager

The Account Manager group collects components related to different user accounts and assigned users' roles. Two main user categories are Investigators and Participants. Investigators use the IoT Lab platform and tools to set-up their investigations or experiments, recruit participants and collect and analyse results. Participants are all the actors involved in an experiment; i.e. people who use the IoT Lab product (application) and allow the experiment execution on their phones. Other users of the platform include: platform owners, researchers/students, testbed owners, customers, testbed service managers, and administrators, etc.

The IoT lab platform will collect different types of data and it will be designed to ensure the privacy and trust of the users. All users will have to be authenticated and appropriately authorised to be able to access the system functionalities. The platform will provide the option for data anonymisation as well as for generating identifiers for such accounts. Access to the platform functionality will be controlled by AAA (Authentication Authorization and Accounting) component. The Security component will control AAA mechanisms in the system. All accounts and roles will be persistently stored in a database.

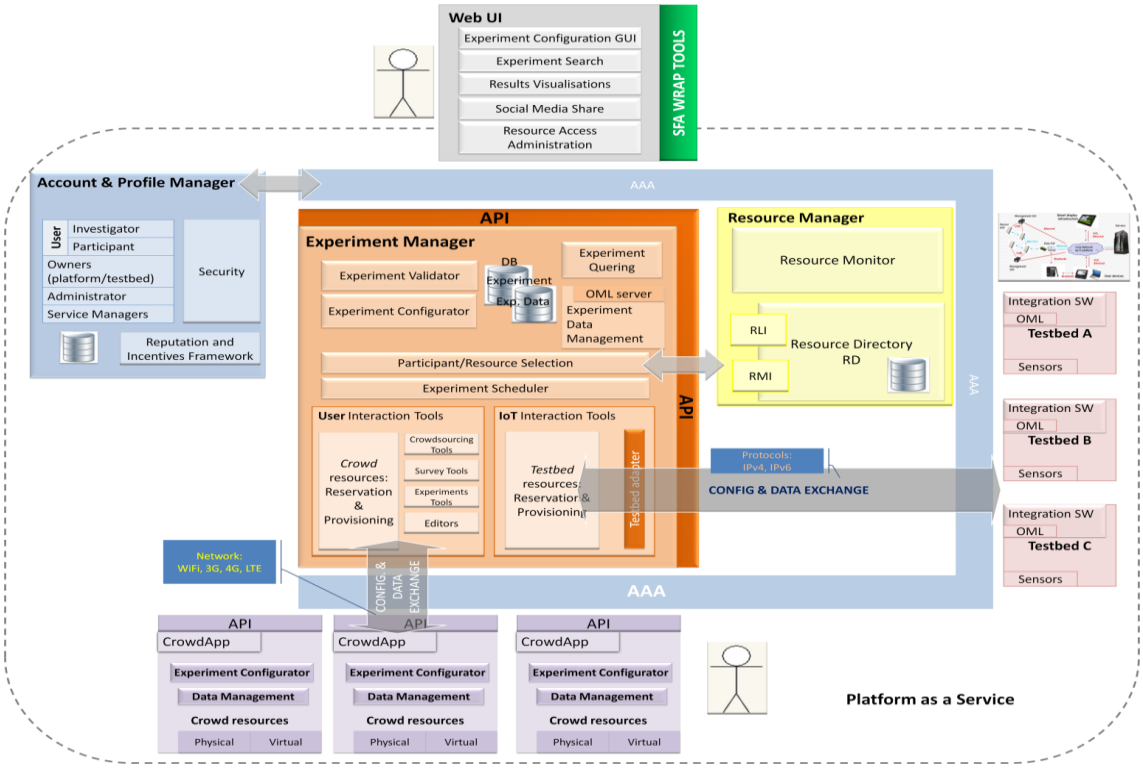


Figure 4 IoT Lab Architecture – A deployment view of the proposed concrete architecture

The function of the Reputation and Incentives framework component is to monitor the users' activity and then estimate the users' rating in a semi-automatic way as well as to apply incentives schemes for engaging different end users and crowdsourcing participants in experiments.

B. *Resource Management*

Resource Management Group monitors and collects information on available resources in the system**.** Resource Directory (RD) component provides a persistent storage of resources. It implements interfaces for the resource management by providing the CRUD functionality as well as the Resource Managing Interface (RMI) and Resource Lookup Interface (RLI) which are implemented as REST based web services. All resources available in the IoT system should be uniformly described. IoT testbed resources management interface uses Fed4FIRE enablers such as Slice-based Facility Architecture (SFA) Wrap and OML to enable interactions with IoT components from testbeds and smart phones in a unified way. SFA Wrap enables resource virtualization, federation and integration of testbeds. OML is responsible for data collection from testbeds and crowdsourced devices.

Resource Monitor Component manages resources in the RD by keeping the real time information on availability of resources in the system.

Testbeds should implement a resource discovery mechanism that will announce their available resources following the RSpec format of the SFA.

C. *Experiment Manager*

Experiment Manager Group aggregates components related to both the experiment management and the experiment data management. An API provides the standardised RESTful interface for component interaction. Several components take part in creating an experiment:

Experiment Validator Component receives a standardised abstract experiment representation and validates the experiment definition. Standardised experiment representation should result from consolidated analysis of use cases and additional user requirements. It can contain code segments that should be performed on the participants' devices involved in the experiment, or a definition of questionnaire forms that should be completed by every participant. All segments of an experiment should be verified and any detected irregularities should be reported before any further processing.

Experiment Configurator Component interprets the received validated experiment definition and stores the standardised experiment representation in an Experiment Database.

Participant/Resource Selection Component detects and selects available resources that match the experiment requirements. This component performs the appropriate query on the RLI interface and receives notifications on availability (and location) of Resources from the Resource Monitor component.

Experiment Scheduler runs the experiment on resources using the Reservation and Provisioning Components for appropriate testbeds.

Experiment Querying, through the Experiment Manager API, provides an access to stored experiments. All data provided by the experiment are collected by the Experiment Data Manager component. Testbeds provide streams of experimental data in the OML format.

Experiments are conducted on top of different testbeds. The process of an experimenter discovering, reserving and provisioning the available resources across all testbeds for his/her experiment will be conducted in a standardised way via the SFA Wrap tools and architecture (e.g. an SFA client will be running at the Web GUI). Then, the IoT Lab Experimenting Platform will take care of the particularities of each testbed and will interact with the resources according to the experiment scenario.

Crowd interaction management interface handles the interaction with the participants. Crowd based experimenting is focused on running the experiments on users' mobile devices. Again, these resources will be exposed by the IoT Lab Experimenting Platform in a standardised way via SFA Wrap.

User Interaction component aggregates components for experiments on top of crowdsourcing smart mobile devices. Several components for survey, crowdsensing and code script execution are involved in experiment execution.

IoT interaction tools control the experiment execution on federated testbeds. All components in the Experiment Manager are accessible through the Experiment manager RESTful API.

D. *Web user interfaces*

GUI access to the system is implemented through components grouped in the Web user interfaces.

The Experiment Configuration GUI provides the Web access for designing and initiating the experiment and it should be intuitive. The GUI communicates with the Experiment Manager through a corresponding API to provide the experiment description to the Experiment Validator. This also includes the access to the Survey and GUI editor so the experimenter can set up a specific survey and/or a specific user interface for his/her experiment on the smart phone application.

The Experiment Search Component provides an interface for the experiment querying. This component can query the resources in order to make the access easier for resources in different testbeds.

Results Visualisation Component provides the appropriate graphical interpretation of collected experimental data. It will include a maps and graphs generator based on main stream open source solutions, such as OGC SWE and Google maps. It will enable the platform to provide graphical representation and dynamic maps of the results as well as of the live data.

Social media will share components of the Web interface and enable different types of users to publish their experiments and related opinions on popular social networks.

The system management is provided through the Platform Administration Component which enables several functionalities including:

- User Account Administration for the users to log in and manage their personal account and profile.
- Data Access & Management for the experimenter to manage the collected data; delete unnecessary data sets; and/or retrieve filtered data sets.
- Filtering users' personal data in order to ensure full compliance with the personal data protection policy and obligations.

V. USE OF SERVICES

An overview of the usage of available services within the proposed IoT Lab platform described in the previous section is provided using the Sequence Diagrams, which illustrate three stages in the platform deployment:

- User registration Process (Fig.5)
- Experiment submission–Investigator side (Fig.6)
- Contribution from Crowd (Fig.7)

A. *User Registration Process*

Testbed Resources: All available individual testbed resources need to be stored in the Resource Manager following their announcement to the IoT Lab platform using a common description scheme. In this way, the IoT Lab platform will be able to leverage the available resources in a uniform manner.

Users of the platform (participants and investigator researchers) should be stored in a Resource Manager component following their registration in Account Manager, authentication through AAA and the role assignment again in Account Manager.

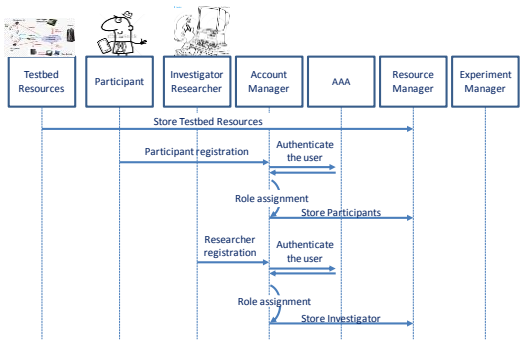


Figure 5: User Registration Process (Sequence Diagram)

B. *Experiment Submission*

Investigator/Researcher uses the Web UI to setup the experiment via Experiment Manager RESTful API.
The experiment is then validated based on standardized experiment representation, interpreted by Experiment Configurator and then stored in an Experiment Database. In addition to automatic validation, experiments are also physically validated as part of a review process involving human resources.

The Experiment Manager is then able to recruit the relevant participants and resources (testbeds) by

communicating the Resource Manager through Resource Lookup Interface (RLI). The experiment is then scheduled by Experiment Manager through reservation and provisioning of resources which includes: Testbeds - reserved through SFA WRAP and Crowd/mobile phones.

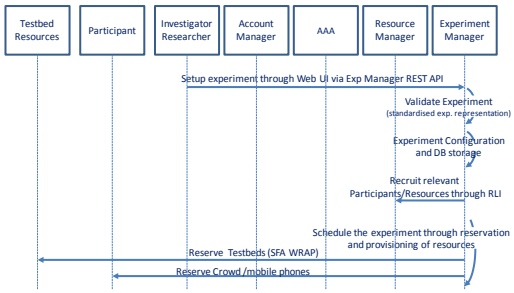


Figure 6 Experiment Submission (Investigator side)

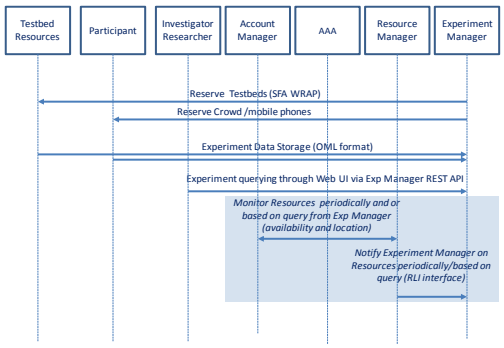


Figure 7 Contribution from the Crowd/Testbed Resources

C. *Contribution from Crowd/ Testbed Resources*

The obtained experimental data from Testbeds and Participants is sent to the Experiment Data Storage within the Experiment Manager in an OML format.

An investigator researcher can send the query for experiments through the Web UI via the Experiment Manager RESTful API.

Monitoring the Resources and information about their availability and location should take place either periodically or based on the query from Experiment Manager between the Account Manager and the Resource Manager.

Resource Manager should notify the Experiment Manager on Resources periodically as well as based on the query through the RLI interface.

VI. CONCLUSIONS

In this paper are proposed the main components of the IoT Lab platform and described their identified functionalities, interaction patterns, interfaces and communication links.

The IoT Lab platform development focuses on:

- Enhancement of existing IoT FIRE testbed facilities which are traditionally built from static sensor mote platforms by including the participants' end user mobile phones and thus achieving a crowd participation in sensing and actuation operations.
- Solutions with built-in reputation and privacy mechanisms as well as procedures for dynamic selection of suitable crowd resources.

The derivation process followed an IoT-A methodology in order to support interoperability and scalability and to enable the use of a wide range of heterogeneous devices as well as the testbeds from different application domains, therefore, satisfying a high number of requirements. Integration of smartphones with existing FIRE testbed infrastructures or any general statically deployed IoT resources represents a novel approach with respect to existing crowdsourcing solutions.

The values that IoT Lab brings in comparison to other existing crowdsourcing solutions are numerous: IoT and crowd sourcing based research and development; access to distributed testbeds; economies of scale; access to innovative service oriented architecture (SOA) technology; a possibility to explore the future now; market insight; crowd interaction and access to wisdom of the crowd; experimental platforms; reduction of development time and time to market; capital expenditure avoidance; cost reduction; access to 'up to date' and evolving IoT services/infrastructures; new product experimentation on advanced IoT facilities.

Our further work will focus on the practical implementation of the proposed platform in order to test selected use cases in real situations and get a feedback from participants. Strategies for engaging as many participants as possible will be examined as well as the new components that provide reward mechanisms towards all participants and investigators.

VII. ACKNOWLEDGMENT

This work has been a part of the collaborative EU-funded project IoT Lab funded from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 610477.

REFERENCES

[1] R. W. Ouyang, A. Srivastava, P. Prabahar, R. Roy Choudhury, M. Addicott, and F. J. McClernon, " If you see something, swipe towards it: crowdsourced event localization using smartphones," in International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), ACM, pp. 23-32, 2013.
[2] Ushahidia. Retrieved from http://www.ushahidi.com/
[3] PhoneLab. Retreived from http://www.phone-lab.org
[4] M. Riahi, T. G. Papaioannou, I. Trummer and K. Aberer. Utility-driven Data Acquisition in Participatory Sensing. 16th International Conference on Extending Database Technology (EDBT), Genoa, Italy, 2013.
[5] V. Agarwal, N. Banerjee, D. Chakraborty, and S. Mittal, "Usense – a smartphone middleware for community sensing," in Mobile Data Management (MDM), 2013 IEEE 14th International Conference on, vol. 1, June 2013, pp. 56–65..
[6] APISENSE. Retrieved from http://www.apisense.com/
[7] EpiCollect. Retrieved from http://www.epicollect.net/
[8] mCrowd. Retrieved from http://crowd.cs.umass.edu/
[9] Funf, Open Sensing Framework. Retrieved from www.funf.org/
[10] Ambient Dynamix. Retrieved from http://ambientdynamix.org
[11] McSense ParticipAct [http://www-lia.deis.unibo.it/Research/McSense/index.html
[12] Internet-of-Things Architecture, http://www.iot-a.eu

Dynamic Software Adapters as Enablers for Sustainable Interoperability Networks

Jose Ferreira*, Carlos Agostinho** and Ricardo Jardim-Goncalves**
* Departamento de Engenharia Electrotecnica, Faculdade de Ciencias e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal. {japf}@uninova.pt
** Centre of Technology and Systems, UNINOVA, 2829-516 Caparica, Portugal. {ca, rg}@uninova.pt

Abstract — **Enterprises are motivated to join collaborative networks, looking forward to reducing time to market, innovating products and lowering prices. However, each enterprise has its own legacy on Enterprise Information Systems, a fact that has been creating a significant interoperability problem when intending to cooperate with others using dissimilar information systems. This paper proposes to reuse existing modelling and architecture technology in a framework to support the sustainability of interoperability among networked enterprises. It suggests the implementation of dynamic software adapters to assure the continuous transformation of heterogeneous information, achieving an adaptive mechanism for interoperation among the different enterprise applications, and guaranteeing a seamless communication among the networked enterprises.**

I. INTRODUCTION

Enterprises cooperate with others to improve their capability to reach new markets, produce with better quality and have cheaper production costs. Hence, globalization brings important opportunities to work with other organizations scattered all over the world. Nevertheless, to achieve seamless collaborative working environments in industrial domains, it is needed an adaptive management of the business processes, supported by software to handle the dynamicity of the business network, formed by the many collaborating partners.

Enterprise Information Systems (EIS) are key towards the management of the enterprise information and it is through them that automatic communication among partnering enterprises can be achieved. EIS are capable of dealing with the use of products and resources (personnel, material, equipment, etc), monitoring operation, or identifying issues such as the delay in production and missing material, which can occur in the enterprise daily activity. Nevertheless, the use of different types of EIS in the same collaborative network may cause problems between the several enterprises within the network [1]. If each uses their own legacy system, the exchange of information becomes more complex since each is following a different information model and a different representation of products and processes. This results in a serious source of interoperability problems', with companies not being able to understand their partner's information. In more difficult situations, the information can even be misunderstood due to semantic divergence, causing production faults that can result in the loss of a considerable amount in the company's profit. Hence, the need for seamless information sharing in a collaborative network is becoming more important [2].

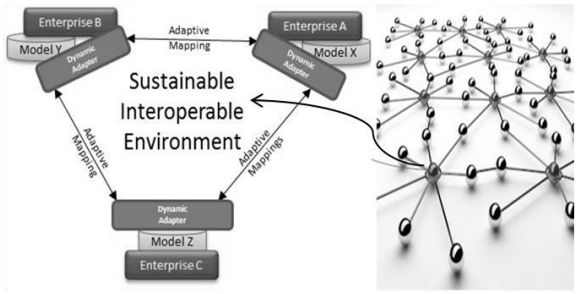


Figure 1. Concept of Sustainable Interoperable Environment

A. Sustainable Interoperability in Networked Enterprises

To evolve further in the advent of the enterprise collaboration, there is the need for Sustainable Interoperability (SI) of the networked environment. SI is defined as "Interoperability that convenes the needs of the present without compromising the ability of future changes, meeting new system requirements, and performing adequate adaptation and suitable management of the transitory elements" [3]. This can be achieved building on the concept of complex systems dynamicity, together with automated methods for adaptive processes and solutions. Indeed, previous research advocates that gradual peer-to-peer mappings, implemented by software adapters, can be established on a need-to-serve basis among different information structures, independently of language, local context, and the number of business relationships within the collaboration networks each enterprise is part of [4]. This paper further develops this idea, investigating how the reuse of existing modelling and architecture technology helps in the semi-automatic generation of software adapters to enable SI.

Figure 1 depicts the concept of a sustainable interoperable environment, detailing a Collaborative Network (CN) instance composed by 3 enterprises cooperating between each other, in an adaptive process. Indeed, a CN is not a static network, and several situations can occur to disestablish the seamless integration of the network (formalized by model mappings). Examples can include the update or change of EIS, new partner with a dissimilar information model, etc. Any of these situations may cause harmonization breaking in the network, which can propagate through the many existing relations (right side of Figure 1. Hence to avoid such situation it is necessary to be prepared to react and adapt in real time to the new circumstances and ensure the interoperability along the whole life cycle of the collaborative environment.

Therefore, the instantiation of the concept of SI Environment benefits from the 3R's principles [5]:

- **Reuse** - Reuse proven methodologies and solutions for EIS specification and modelling, as well as implementation of adaptation and interoperability services. This could avoid big expenses resulting from the need to implement "yet" another solution from scratch;
- **Reduce** - Reduce the time lost on solving interoperability problems. The system will re-adapt using past knowledge and recover the interoperability environment. This will benefit in less loss of money, since time is money;
- **Recycle** - Recycle legacy systems and models, by adapting in-house EIS to new needs. This will avoid drastic changes on the information models and reduce the impact on the network.

Readjust should be added as another principle to the software adapters to ensure the continuous interoperability of heterogeneous and evolving systems. With it, it will be assured an adaptive mechanism for continuous interoperation among the different enterprise applications, and guaranteeing a seamless communication among the networked enterprises. This framework will allow reusing methodologies, reducing costs, recycling legacy systems, and readjusting adaptors, acting in 3 different ways:

1. Create the software adapters to have a first level of interoperability within the network. This level can then be increased along the life-cycle of the collaboration;
2. Maintain the interoperability in the network, monitoring and providing support to dynamicity;
3. Predicting potential problems based on past behaviour at the enterprise level or fluctuations at the network level (companies joining or withdrawing from the network). This will foresee potential issues before they can be a real problem, thus solving them in advance.

During this paper, the presented work is focused in the reuse phase and the creation of the adaptors.

II. ENTERPRISE INFORMATION SYSTEMS INTEROPERABILITY

Enterprise Information Systems (EIS) are integrated application-software packages that use the computational, data storage, and data transmission power of modern Information Technology (IT) to support processes, information flows, reporting, and data analytics within organizations. The integrated content may be used to run a configuration management solution throughout the life cycle in relation to products, assets, processes and requirements of the entity (laboratory, facility, etc.) [6].

EIS are mostly based on commercial software packages (could also be custom developed) without implementing a global strategy for the seamless integration of all the information flowing throughout the company or the network of cooperating enterprises. If an organization has two systems and they cannot seamlessly communicate between each other, this will bring some problems in the productivity and customer responsiveness suffers. In such case, one is facing the problem of fragmented information, and consequently fragmented business [7].

Moreover, EIS use data repositories to collect and store massive amounts of data. Yet, since systems normally evolve with the needs of the companies, and usually different repositories are created, each company has to manage many heterogeneous repositories. This leads the information to be spread across several different computer systems, each hosted in separate business unit, region, factory, or office, without a global seamless capabilities which harnesses efficiency.

From an enterprise architecture and a systems engineering perspective, operating in a networked environment places the requirements for interoperability alongside maintainability, reliability, safety of a system [8]. Hence technologies and frameworks for enterprise modelling are analysed next (non-exhaustively) in order to be reused and support the creation of dynamic software adapters, and the methods necessary for the sustainability of interoperability among networked enterprise.

A. Enterprise Modelling Techniques

Vernadat (2001) defines Enterprise Modelling (EM) as the efficient design, analysis and optimization of enterprise operations requiring notations, formalisms, methods, and tools to depict the various facets of a business organization. A relevant advantage of using EM is that it helps describing the various elements of an enterprise, including its functions, behaviour, information, resources, organization or economic aspects of a given business resource. In order to support enterprises during their modelling process, several methodologies exist. For the purposes of our study a few were selected: Integrated Enterprise Modeling (IEM) [9], GRAI Integrated Methodology (GRAI-GIM) [10] and CIMOSA [11].

CIMOSA has a strong focus on process-oriented approaches aiming at integrating functions by modeling and monitoring the action flow, while GRAI sees integration as the coherence between global and local decision objectives. IEM seeks to support the development of a unified enterprise model and to represent the different aspects of a manufacturing enterprise as views of that model.

Although our study takes greater focus in these referred methodologies, other frameworks exist that are not necessarily less important than the ones here discussed A relevant state of the art of these EA is proposed in [12] and [13].

B. Engineering Modelling Techniques

Each year the need to plan, develop, and manage the enhancements of enterprises infrastructure, products, and services, including marketing strategies for product and service offerings based on new, unexplored, or unforeseen customer needs with clearly differentiated value propositions, is increasing [14]. These events raise the relevance of having engineering modelling within the enterprises themselves. As previously mentioned, in order to better support this modelling process, several methodologies exist that can effectively aid to achieve it. For our purposes, and as already referred previously, this study focused mainly on the following: Model-Based Systems Engineering (MBSE) [15], [16], to formalize modelling in support to the systems engineering processes; Model-Driven Architecture (MDA) [17], [17], to support software engineering involving the multiple actors from the business level down to the programmer; Model Driven

Service Engineering Architecture (MDSEA) [18], [19], merging more classical EM technologies with the engineering perspective of MDA, extending it to the engineering context of product related services in the virtual enterprise environment; and Service-Oriented Modelling Framework (SOMF) [20] for software and web services orientated development.

C. Holistic Modelling Frameworks (Architecture)

Enterprises have the need to internally organize their structures, aiming to be more efficient and achieve their objectives more quickly. For this reason, they seek solutions that are able to guide them in the best course of direction. An example is the creation of an enterprise architecture as an answer for their needs. This is important because an architecture is a framework of principles, guidelines, standards, models, and strategies that direct the design, construction, and deployment of business processes, resources, and information technology throughout the enterprise [21]. To support the enterprises in the modelling several methodologies exist and for our study a few ones were selected: Zachman Framework [22] and TOGAF [23]. They structure various EM and engineering concepts according to the perspectives of various stakeholders involved in the enterprise engineering. Other frameworks exist, however, are not necessarily less important that the ones discussed here [24], [25].

D. Metamodelling

GERAM is a complimentary paradigm that provides a set of recommendations, which are baseline requirements to support enterprise architecture and engineering. This baseline is conceptually located at the meta-modelling level, supporting enterprises in the assessment of the different architectures/methodologies known, and to choose the one that is best suited to their enterprise business needs. Hence, it was developed to encompass and generalize the commonalities of various existing frameworks and reference architectures, by including all knowledge needed for enterprise engineering/integration [26], [27]. It therefore differs from the other architectures/ methodologies, since it just gives support to choose one instead of developing new.

E. Analysis of Different Techniques and Frameworks for EIS specification and Modelling

The several methodologies/frameworks presented are different in nature. Some are more directed to information modelling of the enterprise, others to the information system, while others are more holistic and address the enterprise as a whole, treating requirements, product, process, and software in an integrated manner. All have pros and cons and, in the end, their adoption depends on the cost-benefit, as well as the learning curve for the tools implementing each paradigm.

Figure 2 depicts the coverage of each of the described architectures and methodologies in the operative levels of an EIS, i.e., Business, Process and Service [28]. It illustrates the models and concept used by each paradigm to manage each level of information. From this analysis, it is observed that all addressed architectures and methodologies tackles Process and Service level, whilst MDA, MDSEA, TOGAF, and SOMF additionally embrace the Business level with high level of detail. This means that these methodologies enable interoperability at all levels inside of an organization. In terms nature of data managed, Zachman and TOGAF are both focused on the whole

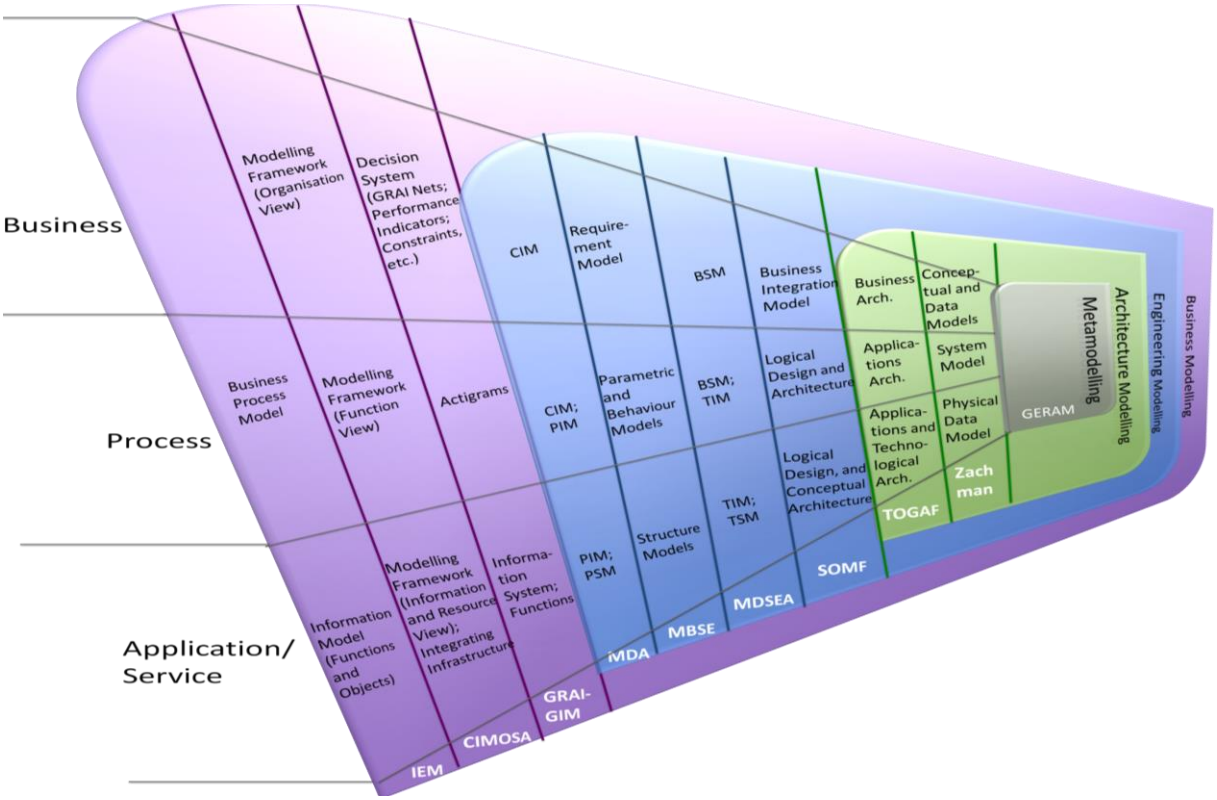


Figure 2. Coverage of modelling architectures and methodologies

enterprise data. CIMOSA and GRAI-GIM are meant to handle computer integrated manufacturing (CIM) data, while IEM is emphasising more on the enterprise process data. MBSE is providing a strong emphasis on the product data and its lifecycle, whilst MDSEA is targeting the manufacturing services. Finally MDA and SOMF are both managing software-related data. In conclusion, all were developed with different purposes, yet, in the end all of them share the same goal, i.e. to provide guidelines for the

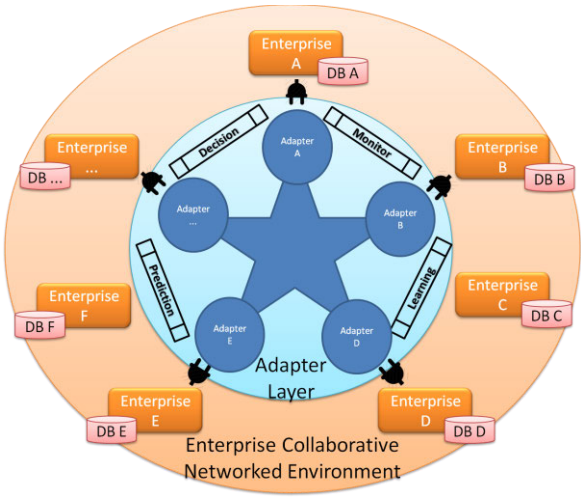


Figure 3. Sustainable Interoperable EIS Network.

enterprises to properly manage their own EIS [8], enabling an interoperable environment inside the enterprise. Yet, none is focused on the network.

III. Towards Sustainable Interoperability

A. Sustainable Interoperability Requirements

As explained before, the sustainable interoperability is an ideal to allow enterprises to adapt their EIS and still maintain the interoperability along its operating life cycle. However, to reach this status is needed to identify requirements that will give the possibility to fulfil the desired goal. To have a sustainable interoperable environment, it is important to understand the modelling paradigm in use, the relationship among the business partners so that intelligent reconfiguration of components becomes possible. To create this maintenance system and manage such dynamics, it is necessary to **monitor** and adapt to the changes while **learning** over time. Nevertheless, in another perspective it is needed to **predict** the transient that results from the dynamics of the individual systems, since a network (and network of networks) will face changes that impact third parties in the global operative environment. Hence, an evolution of a particular system should only be **decided** in case it brings more benefits than damages. According to the reflectivity principle [29], changes can follow a cyclic loop that impacts the same system that motivated the initial evolution.

Another important aspect is to have a conformance test and interoperability checking for systems interoperability assessment. This assessment is needed to discover and notify every time that a new system node is integrated in the collaborative network, or it is updated. The conformance checking is required to check for conformance of data, models, knowledge and behaviours of the systems and assure accuracy in the seamless communication. The interoperability checking will verify and assess the network to assure the maintenance of the network interoperability system [30], [31].

B. Dynamic Software Adapters at Work

To exemplify the concept, let's consider a collaborative network composed by several enterprises (nodes A, B, C, etc.), as represented in Figure 3. To create an interoperable environment on this collaborative network, it is necessary to identify the models of each enterprise, to establish and assure the integration between each other. From this circumstance it can happen different situations in the network: Company A uses the same EIS as Company B (in this case the integration should be easier); Company C has an EIS that is different from the one owned by Company A and B, but still, it has interoperability potential. The software adapter is the component responsible for the identification of the models and mapping between each other, that as the name implies, has the function of adapting each model to another. Creating a way to propose a solution to do the relations between them is an important step to achieve the global interoperability status in the network. Due to this, the study made in section 2 is important to identify how the EIS work, and how they represent their models. Since allowing the different enterprises to use their preferred systems is important, the adapter is reusing the legacy systems of each enterprise, creating a sustainable environment in the network.

In Figure 3, one can see how each enterprise uses its specific adapter to communicate with others, and are associated with the monitoring, learning, prediction and decision activities, directly related with the systems evolution and adaptation.

C. Methodology for the Semi-Automatic Generation of Software Adapters reusing on Enterprise Models

A crucial characteristic to be considered in the support for the sustainable interoperability concept, is to allow the network to detect the changes that occur over time, and adapt to them using semi-automatic technology such as

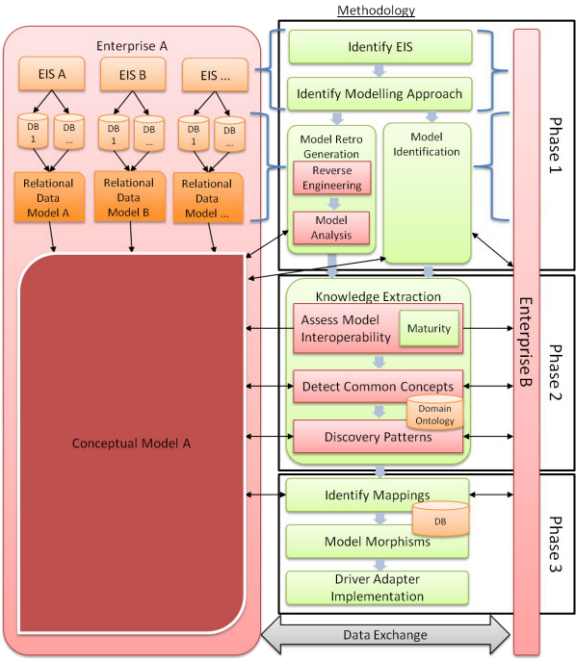


Figure 4. Methodology proposed.

software adapters to translate data from one system to the other, giving the possibility to learn, as well as to provide the capabilities for monitoring, supervision, and context awareness. The authors propose a methodology for the semi-automatic generation of software adapters, applying and reusing existing enterprise data, models and knowledge from modelling paradigms such as the ones analyses in section 2.

As illustrated in Figure 4, the methodology details the process followed to put two or more enterprise information systems exchanging the information needed by the partner enterprises (A and B) to collaborate. It is divided in three phases as described next:

Phase 1 - A first step (manual) is related with the identification of all the EIS in use. As each enterprise can have different systems, e.g. ERP, MES, CRM, etc., it is important to start by identifying which are them, and which are relevant for the collaboration. Directly related with this, is the identification of the modelling approach (if any). As discussed along section 2, enterprise modelling can follow different paradigms, maintaining different types of models and information views. At this point one can immediately access the conceptual models needed, or apply an automated retro-generation procedure to inspect the local databases and derive one or more relational models (see the work of Lezoche et al. [32]), which can then be abstracted to conceptual level. The purpose of these initial steps is to identify the conceptual data models that need to be interoperable.

Phase 2 - The second phase of the methodology is related with knowledge extraction to initially assess the models interoperability status and/or requirement, detect the common or similar concepts in use (with the help of domain ontologies), and discover patterns of existing relationships. These semi-automatic steps enable to identify the mappings needed to relate both enterprises. If there is already an adapter implemented to relate both enterprises, this phase will also validate the existing mappings if they are stored explicitly (as in the work of Agostinho et al. [33]).

Phase 3 – Having the mappings duly identified, one can apply the software engineering modelling paradigms of section 2, e.g. MDA, to formalize the morphisms using models and enable semi-automated implementation of the transformation services/functions based. This phase is not so thoroughly detailed here because there is parallel research exclusively focused on the topic (e.g. Agostinho et at. [4]).

After the implementation of the software adapter, the information can be exchanged as needed, and there will be enough meta-information on the adaptor to enable to regenerate it based on new modelled morphisms, whenever necessary for a sustainable interoperability. The methodology is in an experimental phase. Yet, some parts were already tested and validated. In the Phase 1 the authors are making tests in the Model Retro Generation module, where they are researching for methods to improve the Model Analysis used in the IMAGINE project. The Phase 3 is the phase that have more results, the Identify Mappings and Model Morphims modules were already validated and it is being developed a way of auto generate automatically the transformation of the models, allowing the system to readapting the changes in the models without breaking the harmonization in the network.

D. Adapter Implementation

The adapter implemented following the methodology proposes is able to identify the EIS, models and standards in use by the enterprise, and to use that new knowledge to create a bridge that will allow the integration between other enterprises. In terms of architecture, the adapter will be composed of different modules that contribute with different targets such as monitoring, decision, etc., and reach the final goal, establishing an SI environment and avoiding harmonization breaking in the network.

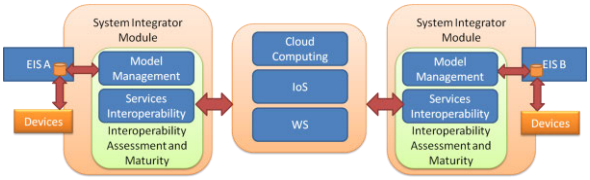


Figure 5. System Integration Module (SIM).

In this paper only a part of the adapter is addressed, namely the System Integration Module (SIM), as illustrated in Figure 5, and whose model and services for interoperability results from the presented methodology. Model Management component is used to display and relate the different models, enabling an interaction with the user that can complement the results of the automatic mattings obtained. The Service Interoperability component creates the physical bridges in the communication, relation with other SIM modules from other adapters through cloud technology. With these components, the adapter is ready to evaluate and relate the different enterprise models, and giving support to the three R's more exactly to the Reuse, since it is reusing proven methodologies and the EIS conceptual models used by each enterprise even they are not explicitly defined at the beginning.

To validate the methodology and the module presented, it was developed a first approach in the IMAGINE project [34]. A software adapter, following the same principles discussed here was implemented for a specific cluster of companies to connect to a services platform (refer to for specific details on the implementation conducted there [35]). However, due to project requirements, the approach was made in a different perspective, since this methodology proposes a decentralized approach (each company has its specific adapter) as opposed to the centralized implementation from IMAGINE, where there is a single complex adapter for a full set of enterprises. Since, in that case an enterprise only needed to create mappings to transform their data to the model of the platform (unidirectional), that approach was possible. The adapter developed for the IMAGINE project is represented in Fig. 3, where each enterprise connects to the adapter to provide the needed information data to be used by the IMAGINE platform. Each adapter is specific for each enterprise, since their models are different. So, it was needed to create specific mappings to allow the flow of information between enterprises and platform.

In conclusion, during the IMAGINE implementation, phase one of the methodology (retro-generation path) was applied to several EIS, phase two was skipped since the modelling of phase three was done manually due to the degree of dissimilarity of the destination model. Some parts were manual/semi-automatic (as the identification of the model and identification of the mappings) and others parts were automatic (as the transformation of the models),

making the first approach of the adapter with good results. In the adapter of the project, it was missing an automatically system to identify changes in the models, allowing the adapter to identify and to do the changes in real-time without creating a harmonization breaking in the network. This is a point needed to allow the methodology to be dynamic, being a point for future work and in study as a result for this work.

IV. Conclusions and Future Work

Adaptive systems theory is taken as a basis for this work, considering that EIS collaborative networks are a complex macroscopic collection' of relatively similar and partially connected micro-structures, formed in order to adapt to the changing environment and increase its survivability as a macro-structure. It is particularly difficult to maintain the interoperability in such enterprise collaborative network. This is mainly because of how EIS systems are designed, since each system usually has different information models and interfaces.

To assure the sustainability of interoperability in a network of EIS, it is necessary to create seamless relations between the heterogeneous systems, its standards and models. In this paper, it was made a study on several EIS modelling paradigms, to identify how they work and how they represent the enterprises and their own resources, enabling the reuse of proven methodologies and systems' models in a semi-automatic generation of software adapters to enable SI. If every time enterprises need to cooperate, the philosophy of the three R's is applied, it reduces costs, improves the efficiency, and maximizes the collaboration potential of networks.

This paper proposed a paradigm to support the sustainability of interoperability among networked enterprise, suggesting the implementation of dynamic adapters to assure the continuous transformation of heterogeneous information models. Following a 3 phase's methodology, it assures an adaptive mechanism for continuous interoperation among the different enterprise applications, and guaranteeing a seamless communication among the networked enterprises. To achieve this sustainability, it is necessary to identify the EIS/models in use, extract knowledge out of them to enable automation in the creation the mappings between the different models. An architecture for the software adaptors has been proposed and implemented in a trial case for the IMAGINE European project.

In the future, the knowledge extraction part of the methodology is to be further developed, creating a solid basis for automatic reasoning. Also, the management of such complexity and number of relations requires an evolved model management human-interface, supporting decision making and enabling simulation of transients in the different networks each enterprise belongs to. Complex Event Processing (CEP) and MAS (Multi-Agent System) are examples of candidate technology that may improve the monitoring of networks.

Acknowledgment

The research leading to these results has received funding from the European Union 7th Framework Programme (FP7/2007-2013) and Horizon2020 (H2020/2015-2020) under grant agreement: IMAGINE (FP7-285132) (www.imagine-futurefactory.eu/), C2NET (FoF-01-2014 nº636909) and OSMOSE (FP7-610905).

References

[1] ATHENA Project, "D.B3.1 - Business Interoperability Framework," 2007.

[2] R. Jardim-Goncalves, C. Agostinho, J. J. Sarraipa, A. R. de Togores, M. J. Nuñez, H. H. Panetto, and M. J. NuneZ, *Standards Framework for Intelligent Manufacturing Systems Supply Chain*. InTech, 2011.

[3] C. Agostinho and R. Jardim-Goncalves, "Dynamic Business Networks: A Headache for Sustainable Systems Interoperability," *4th Int. Work. Enterp. Integr. Interoperability Netw. OTM Conf.*, 2009.

[4] C. Agostinho, P. Pinto, and R. Jardim-goncalves, "Dynamic Adaptors to Support Model-Driven Interoperability and Enhance Sensing Enterprise Networks," in *The 19th World Congress of the International Federation of Automatic Control, Cape Town, South Africa*, 2014.

[5] NRDC, "The 3R's Still Rule," 2008. [Online]. Available: http://www.nrdc.org/thisgreenlife/0802.asp.

[6] US Departement of Energy, "DOE Standard - Content of System Design Descriptions," 2011.

[7] T. H. Davenport, "Putting the enterprise into the enterprise system.," *Harv. Bus. Rev.*, vol. 76, no. 4, pp. 121–31, 1998.

[8] P. Kotze and I. Neaga, "Towards an Enterprise Interoperability Framework," in *Advanced Enterprise Architecture and Repositories and Recent Trends in SOA Based Information Systems: In Conjunction with ICEIS 2010*, 2010.

[9] M. Kamath, N. Dalal, A. Chaugule, E. Sivaraman, and W. Kolarik, *Scalable Enterprise Systems: An Introduction to Recent Advances*. 2003.

[10] J. M. c. Reid and N. D. Preez, "Evaluation of the Grai Integrated Methodology and the IMAGIM Supportware," *South African J. Ind. Eng.*, vol. 9, no. 1, 1998.

[11] PERA, "CIMOSA - CIM Open System Architecture," *PERA Enterprise Integration Web Site*, 2001. [Online]. Available: http://www.pera.net/Methodologies/Cimosa/CIMOSA.html. [Accessed: 14-Sep-2013].

[12] CHEN D., D. G., and VERNADAT F., "Architectures for Enterprise Integration and Interoperability: Past, present and future," *Comput. Ind.*, vol. 29, no. 3, pp. 647–659, 2008.

[13] V. F., "Enterprise Integration and Interoperability," in *Handbook of Automation*, Springer-Verlag, Ed. Berlin, 2009, pp. 1529–1538.

[14] R. Pineda, A. Lopes, B. Tseng, and O. Salcedo, "Service Systems Engineering: Emerging Skills and Tools," *J. Procedia Comput. Sci.*, vol. 8, pp. 420–427, 2005.

[15] T. Operations and H. Crisp II, "Systems Engineering Vision 2020," *INCOSE*, vol. Version 2., no. September, 2007.

[16] INCOSE, "INCOSE," 2011. [Online]. Available: www.incose.org.

[17] A. Petzmann, M. Puncochar, C. Kuplich, and D. Orensanz, "Applying MDA ® Concepts to Business Process Management," *Interchange*, pp. 103–116, 2007.

[18] MSEE Project, "D11.2 Service concepts, models and method: Model Driven Service Engineering," 2012.

[19] H. Bazoun, G. Zacharewicz, Y. Ducq, and H. Boye, "Transformation of Extended Actigram Star to BPMN2.0 and Simulation Model in the frame of Model Driven Service Engineering Architecture," in *DEVS'13 - Proceeding of the Symposium on Theory of Modeling & Simulation*, 2013.

[20] Methodologies Corporation and M. Corporation, "Service-Oriented Modeling Framework (SOMF)," 2011. [Online]. Available: http://www.modelingconcepts.com/pages/SOMF_IMG.html. [Accessed: 24-Sep-2013].

[21] Technology Training Limited, "Introduction to Enterprise Architecture," 2013. [Online]. Available: http://www.technology-training.co.uk/introductiontoenterprisearchitecture_30.php. [Accessed: 16-Sep-2013].

[22] J. Zachman, "The Framework for Enterprise Architecture: Background, description and utility, Zachman Institute for

Advancement," 1996. [Online]. Available: http://www.zifa.com.
[23] A. Josey and The Open Group, "TOGAF Version 9.1 Enterprise Edition: An Introduction." 2011.
[24] K. Mertins and R. Jochem, "Architectures, methods and tools for enterprise engineering," *Int. J. Prod. Econ.*, vol. 98, no. 2, pp. 179–188, 2004.
[25] L. Urbaczewsk and S. Mrdalj, "A Comparison of Enterprise Architecture Frameworks," *Inf. Syst.*, vol. VII–2, no. 18, 2006.
[26] P. Saha, "Analyzing The Open Group Architecture Framework from the GERAM Perspective." 2004.
[27] IFIP-IFAC Task Force, "GERAM: Generalised Enterprise Reference Architecture and Methodology." 1999.
[28] ISO, "Advanced automation technologies and their applications — Part 1 : Framework for enterprise interoperability (ISO/DIS 11354-1:2011)," 2011.
[29] E. Honour, "Systems Engineering and Complexity," *INCOSE Insight*, vol. 11, no. 1, p. 20, 2008.
[30] R. Jardim-Goncalves, C. Agostinho, and A. Steiger-Garcao, "A Reference Model for Sustainable Interoperability in Networked Enterprises: Towards the Foundation of EI Science Base," *Int. J. Comput. Integr. Manuf.*, vol. 25, 2012.
[31] F. Vernadat, "UEML: Towards a Unified Enterprise Modelling Language," in *3º Conference Francophone de MOdelisation et SIMulation "Conception, Analyse, et Gestion des Systemes Industriels", MOSIM'01*, 2001.
[32] M. Lezoche, A. Aubry, and H. Panetto, "Formal Fact-Oriented Model Transformations for Cooperative Information Systems Semantic Conceptualisation," in *Enterprise Information Systems Lecture Notes in Business Information Processing*, vol. 102, Springer, 2012, pp. 117–131.
[33] C. Agostinho, J. Sarraipa, D. Goncalves, and R. Jardim-Goncalves, "Tuple-Based Semantic and Structural Mapping for a Sustainable Interoperability," in *Ifip International Federation For Information Processing*, 2011, pp. 45–56.
[34] IMAGINE, "IMAGINE," 2011. [Online]. Available: http://www.imagine-futurefactory.eu/index.dlg.
[35] J. Ferreira, F. Gigante, J. Sarraipa, M. J. Nunez, C. Agostinho, and R. Jardim-Goncalves, "Collaborative Production using Dynamic Manufacturing Networks for SME's," in *20th ICE - International Conference on Engineering, Technology and Innovation - IEEE TMC Europe Conference*, 2014.

Smartphone MEMS Accelerometer for Cycling – Observations

Sara Stančin, Sašo Tomažič
University of Ljubljana, Faculty of electrical engineering, Ljubljana, Slovenia
sara.stancin@fe.uni-lj.si, saso.tomazic@fe.uni-lj.si

Abstract— **Equipped with 3D MEMS sensors, today's smartphones provide for a wide range of practical measurements. However, MEMS sensors are characterized by different errors including zero level offset, inaccurate sensitivity, axis misalignment and noise. Bearing these inaccuracies in mind, we investigate the feasibility of using a smartphone 3D accelerometer during cycling and provide with some practical observations. Using the accelerometer, we track the slope of the road. Binding the obtained road slope profile with the GPS speed measurements, we estimate motion altitude change. We conclude that, if suitably calibrated, the accelerometer can be feasible for this purpose. However, even small errors in gravity projections estimates lead to significant errors in altitude change estimation. Due to integration, measurement noise contributes less to the estimated altitude error.**

I. INTRODUCTION

A 3D accelerometer enables measurements of acceleration caused by gravity and self-accelerated motion along the three orthogonal sensitivity axes. As such, accelerometers are increasingly being used for simple and frequent human motion measurements enabling human motion capture [1, 2], monitoring [3, 4], analysis and characterization [5-10]. Different support for rehabilitation [11-13] as well as for augmented and virtual reality [14, 15] has been proposed.

Today available kinematic sensors that are based on Microelectromechanical systems (MEMS) are small, light, widely affordable, and come with their own battery supply. These sensors cause minimal physical obstacles for motion performance and can provide simple, repeatable, and collectible motion data indoors. Moreover, because of their low energy consumption, MEMS sensors are a promising tool for tracking motion outdoors.

Widely available low-cost MEMS sensors are characterized by different types of sensor inaccuracies (including the zero level offset, inaccurate sensitivity and misalignment of the sensor sensitivity axes), temperature drift and noise. If non-negligible acceleration is present in the motion, the typical obstacle is the correct deduction of the gravitational acceleration from the total measured acceleration. Because the position data are obtained by integrating the acceleration twice, even small errors in the determined direction of acceleration can cause the calculated position to deviate considerably from the true sensor position. Due to the accumulation of position error, efficient implementation of navigation applications using low cost MEMS accelerometers and gyroscopes alone is not possible.

In this research, we investigate the feasibility of using a smartphone 3D accelerometer during cycling. We focus on the road slope and altitude change estimation.

By providing for load monitoring and energy management, these data are very valuable to the average recreational or professional cyclist.

The result of the accelerometer sensitivity to gravity is that when at rest, the accelerometer shows 1 g of acceleration along the axis of sensitivity oriented along the direction normal to the horizontal surface. This makes it easy to determine the orientation with respect to the direction of the vector of gravitational acceleration in the accelerometer coordinate system.

It is a necessary condition that the accelerometer is either stationary either moving with negligible acceleration in relation to the gravitational acceleration either that the speed of the accelerometer is also by some mean known.

For our purpose, we focus on the first scenario and track bicycle motion characterized by negligible acceleration in relation to the gravitational acceleration. In such a way, the projections of gravity onto the sensor sensitivity axes directly reflect the slope of the road.

Further on, by binding the accelerometer measurements with speed data, we estimate motion altitude change.

II. BICYCLE MEASUREMENT SETUP

To investigate the feasibility of a smartphone accelerometer for tracking road slope and altitude change during cycling, we mounted an iPhone 4S on the handlebar of a bicycle as shown in Figure 1.

The orientation of the smartphone accelerometer intrinsic coordinate system is as illustrated in Figure 1. φ_0 refers to the angle between the sensor and the direction of motion. In further text, we will refer to this angle as the accelerometer level angle.

The smartphone accelerometer is designed in such a way that when at rest it measures 1g of acceleration along the downwards pointed axis.

In this experiment, we relied on the GPS system for the speed measurements. The GPS receiver position data errors are in 10 m range. Speed obtained by tracking position change is for this reason highly unreliable. However, the GPS also provides for speed measurements based on the Doppler effect. For our purpose, these have shown to be much more reliable.

To capture both, the accelerometer and GPS data at common times samples, we used the iPhone SensorLog application. The measurement range of the accelerometer is set to ± 2g and cannot be changed.

An alternative for obtaining the reference speed is using a dedicated bicycle speed measurement unit. Such a unit tracks wheel rpm and in such a way provides for onsite speed measurements. In such a measurement setup, the user would have to provide for efficient synchronization of the accelerometer and the speed measurement unit.

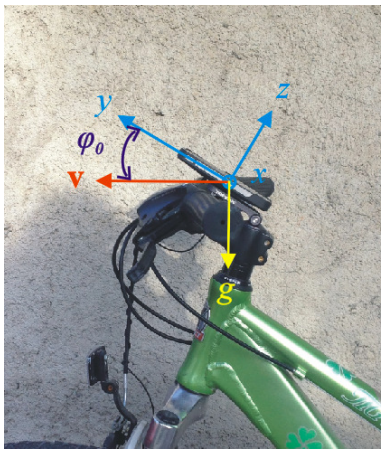


Figure 1. Side view of the smartphone position on the handlebar of the bicycle.

III. ACCELEROMETER CALIBRATION

When analyzing motion dynamics, just as in the case of absolute motion values estimation, accurate data are the basis for an effective and a comprehensive analysis. The accuracy of the captured data is essential for relevant and comparable results. The first step in motion data capture and analysis is hence sensor calibration.

According to the generally adapted model, the accuracy of the values measured with a 3D sensor is influenced by the accuracy of the sensor axis sensitivity, zero level offset and orientation.

The aim of different calibration procedures is to compensate for the measurement errors that arise because of the enlisted inaccuracies.

Considering static calibration, we need to obtain values of 12 calibration parameters. 3 of these are needed to compensate for zero level offset and the remaining nine compensate for axes misalignment and inaccurate sensitivity.

Different procedures have been proposed in order to achieve sensor calibration. We performed the 3D accelerometer calibration according to the procedure we have already presented in [16]. This procedure provides for, in terms of lifetime and computational complexity, an efficient calibration procedure that does not require any additional expensive equipment and is suitable for everyday practical use.

The procedures exploit the fact that the value of the measured acceleration at rest is constant and equal to gravity acceleration. To estimate the values of the 12 calibration parameters contained, six measurements are needed. During these six measurements, the sensor is at rest on an even horizontal surface in turn in six different orientations. The orientations of the first three measurements are such that the sensor intrinsic axes *x*, *y*, and *z* show in turn in the direction of the gravity vector **g**. The orientations of remaining three measurements are such that the sensor intrinsic axes *x*, *y*, and *z* show in turn in the opposite direction of the gravity vector **g**.

For each of the six measurements, we obtained 10,000 samples. The obtained data are shown in Figures 2 and 3.

From these figures we can observe that each of the three sensor intrinsic sensitivity axis (*x*, *y* and *z*) has a negative zero level offset which is the greatest for the *z*-axis. The measurements are not symmetrical for pairs of measurements (*x*/-*x*, *y*/-*y* and *z*/-*z*) from what we can conclude that the axis are not perfectly aligned.

Averaging values detected for each of the six measurements, matrices $\mathbf{A}_{s+}$ and $\mathbf{A}_{s-}$ were calculated. The columns of both of these matrices are equal to the vectors of the detected acceleration values for each of the three measurements, and the rows represent the sensitivity axes. Matrix $\mathbf{A}_{s+}$ combines the three measurements when the three sensor coordinate axes *x*, *y* and *z* is, in turn, aligned with the direction of **g**. Matrix $\mathbf{A}_{s-}$ combines the three measurements when the three sensor coordinate axes are in turn, aligned with the direction opposite to the direction of **g**.

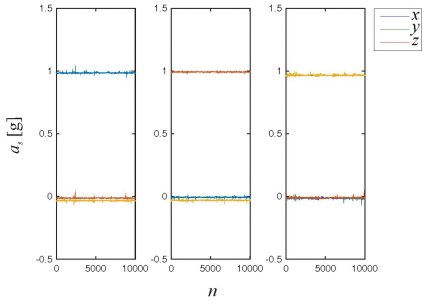


Figure 2. IPhone 4s 3D accelerometer collected data for the first three calibration measurements. During these measurements, the sensor orientations were set such that the sensor's intrinsic coordinate axes (*x*, *y* and *z*) were aligned in turn along the direction of the gravity vector **g**.

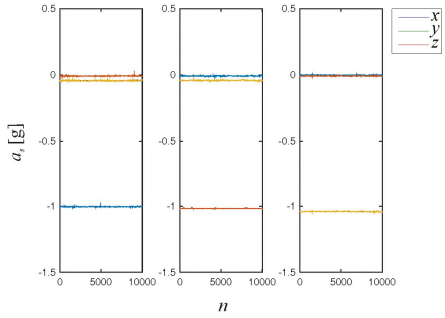


Figure 3. IPhone 4s 3D accelerometer collected data for the remaining three calibration measurements. During these three measurements, the sensor orientations were set such that the sensor's intrinsic coordinate axes (*x*, *y* and *z*) were aligned in turn opposite to the direction of the gravity vector **g**.

Using matrices $\mathbf{A}_{s+}$ and $\mathbf{A}_{s-}$ we obtain the zero level offset vector and the calibration matrix according to [16]:

$$\mathbf{a}_o = \frac{\left(\mathbf{A}_{s+} + \mathbf{A}_{s-}\right)\times\mathbf{i}}{6} \qquad (1)$$

$$\mathbf{C}_s = 2\left(\mathbf{A}_{s+} - \mathbf{A}_{s-}\right)^{-1} \qquad (2)$$

Values obtained for the used iPhone 4S device were:

$$\mathbf{a}_o = \begin{bmatrix} -0.0066 \\ -0.0094 \\ -0.0358 \end{bmatrix} \quad (3)$$

$$\mathbf{C}_s = \begin{bmatrix} 1.0058 & -0.0004 & 0.0067 \\ 0.0029 & 0.9958 & 0.0013 \\ -0.0050 & -0.0050 & 0.9969 \end{bmatrix} \quad (4)$$

Each acceleration measurement triplet $\mathbf{a}_s$ is than corrected according to:

$$\mathbf{a} = \mathbf{C}_s \times (\mathbf{a}_s - \mathbf{a}_o) \quad (5)$$

From the zero level offset results (3), we can conclude that, as expected, the zero level offset is the greatest for the *z*-axis.

From the obtained calibration matrix (4), we can estimate the vector of accelerometer sensitivities **S**:

$$\mathbf{S} = \begin{bmatrix} 0.9942 \\ 1.0042 \\ 1.0031 \end{bmatrix} \quad (6)$$

where each vector element represents the sensitivity of the respective axis for the used accelerometer.

From the obtained calibration matrix (4), we can also estimate the sensitivity axes alignment matrix:

$$\mathbf{\Psi} = \begin{bmatrix} 0.3857 & 89.9789 & 90.3851 \\ 90.1656 & 0.1812 & 90.0736 \\ 89.7160 & 89.7122 & 0.4043 \end{bmatrix}. \quad (7)$$

In the above matrix, all angles are given in degrees. The rows of matrix **Ψ** represent the sensor sensitivity axes, and its columns represent the sensor coordinate axes. From (7) we can conclude that among the three axes of the used sensor, *y*-axis has the least misalignment.

IV. MEASUREMENT ANALYSIS

A Measurement route

The observed cycling route included a hill climb on an asphalt road with a total altitude change of 506 m. The whole route lasted 1 h.

B Measurement Data

The obtained measurement data was sampled in uneven sampling intervals ranging from 6 ms to 1.617 s. For this reason, both accelerometer and GPS measurements were interpolated to the shortest sampling interval from the original data. In such a way, we obtained data sampled at f_s = 1/6ms = 166.7 Hz.

Data obtained for all three sensor axes are illustrated in Figure 4.

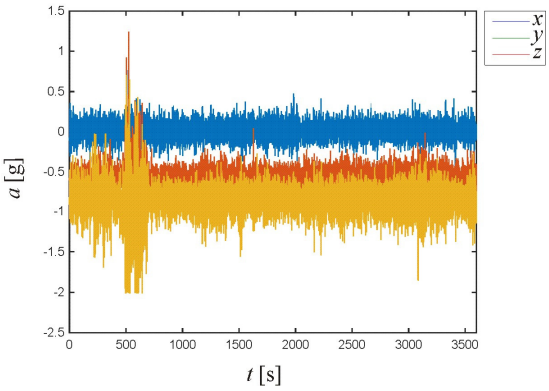


Figure 4. Measurement 3D acceleration data after calibration.

In further analysis, we made the following assumptions:

1. We supposed that the bicycle acceleration is neglectable in comparison to gravity. When cycling uphill, this is a reasonable assumption. This enables us to obtain the slope of the road considering the projection of gravity onto axes *y* and *z*. However, from Figure 4, we can observe high oscillations of the measured acceleration along all three axes that are a consequence of the roughness of the asphalt road. Due to the limited measurement range, these oscillations cause clipping of the measured data along the *z*-axis. For this reason, we could not rely on the *z*-axis acceleration data.
2. For our investigation, we simplified the problem and supposed that the lateral tilt of the bicycle is not affecting the gravity projection onto *y*-axis.

Under these assumptions, we can estimate the road slope considering only the acceleration measured along the *y*-axis. Considering the orientation of the sensor axes as illustrated in Figure 1, we can write:

$$\varphi = \cos^{-1} a_y - 90^\circ - \varphi_0 \quad (8)$$

where φ_0 is the accelerometer level angle (i.e., the angle between the accelerometer *y*-axis and the direction of motion - see Figure 1). This angle can be roughly estimated prior to measurements, from the a_y measurements at an approximately level surface.

Before actual road slope calculation, we filtered the re-sampled calibrated accelerometer outputs with a low pass filter. The cutoff frequency was chosen to be f_{co}=0.14 Hz. As estimated, choosing this frequency provided for the best distinction between vibrations due to the roughness of the asphalt road and the true slope of the road.

C Results

Using the road slope and equally re-sampled GPS speed data, we can calculate the change in the bicycle altitude:

$$\Delta h = \sin(\varphi)\, v \Delta t \quad (9)$$

where *v* represents bicycle speed obtained from the GPS, Δt is the time interval and φ represents the road slope determined according to (6). The actual angle φ_0 in (6) is estimated by finding the best fit solution for height change determined according to (7) and GPS altitude data. The

best fit was obtained for φ_0=31˚. The resulting road slope φ throughout the experiment route is shown in Figure 6.

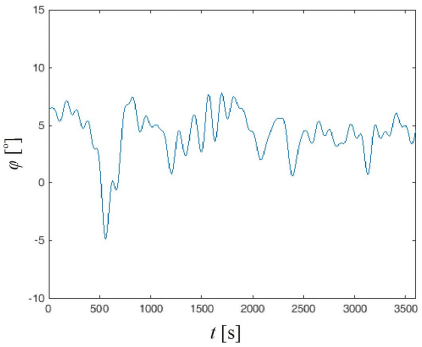


Figure 5. Road slope obtained from the smartphone accelerometer.

For result evaluation, we sum the altitude changes obtained according to (7) and compare them to GPS obtained altitude data. Altitude results for φ_0=31˚ are shown in Figure 6. GPS altitude data is shown as reference data. We can observe that the results match closely throughout the 1 h ride. The figure also shows the resulting altitude data obtained considering a small deviation in the level angle of the smartphone φ_0=31.02˚. At the beginning of the route, the accelerometer altitude data follows the GPS altitude. However, after a 1h ride, an error of 0.02˚ in the estimated level angle causes a 30.84 m altitude estimate error.

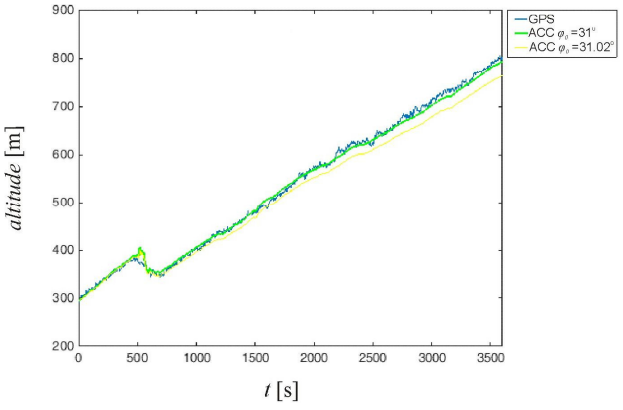


Figure 6. Altitude obtained from the smartphone accelerometer road slope data combined with GPS speed data.

V. CONCLUSION

From the results obtained, we can conclude that, if suitably calibrated, the smartphone accelerometer can be used for road slope tracking and altitude change estimate. In this article we presented a way of performing this task when the motion of the bicycle is characterized by negligible acceleration in relation to the gravitational acceleration. In such a way, the projections of gravity onto the sensor sensitivity axes directly reflect the slope of the road. Considering altitude change estimates, we can conclude that due to integration, even small errors in the estimated gravity projection cause immense errors in altitude estimation. Further on, integrating measurement noise has a low pass filtering effect and for this reason noise itself contributed less to the error in the estimated altitude.

To achieve road slope and altitude estimation in general, highly accurate bicycle velocity data would be needed.

REFERENCES

[1] Aminian, K., Najafi, B., Capturing human motion using body-fixed sensors: outdoor measurement and clinical application, Comp. Anim. Virtual Worlds, 15, pp. 79–94, 2004.

[2] Roetenberg, D., Slycke, P.J., Veltink, P.H., Ambulatory Position and Orientation Tracking Fusing Magnetic and Inertial Sensing, IEEE Trans. Biomed. Eng., 54(5), pp. 883–890, 2007.

[3] Aminian, K., Robert, P., Buchser, E.E., Rutschmann, B., Hayoz, D., Depairon, M., Physical activity monitoring based on accelerometry: validation and comparison with video observation, Med. Biol. Eng. Comput., 37, pp. 304–308, 1999.

[4] Uiterwaal, M., Glerum, E.B.C., Busser, H.J., Lummel, R.C., Ambulatory monitoring of physical activity in working situations, a validation study, J. Med. Eng. Tech., 22(4), pp. 168–172, 1998.

[5] McIlwraith, D., Pansiot, J., Yang, G.Z., Wearable and ambient sensor fusion for the characterisation of human motion, 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5505–5510, 2010.

[6] Heinz, E.A., Kunze, K.S., Gruber, M., Bannach, D., Lukowicz, P., Using Wearable Sensors for Real-Time Recognition Tasks in Games of Martial Arts - An Initial Experiment, 2006 IEEE Symposium on Computational Intelligence and Games, pp. 98–102, 2006.

[7] Hoffman, M., Varcholik, P., LaViola, J.J., Breaking the status quo: Improving 3D Gesture Recognition with Spatially Convenient Input Devices, Proceedings of the 2010 IEEE Virtual Reality Conference (VR); Waltham, MA, USA. 20–24 March 2010, pp. 59–66. 2010.

[8] Lin, P.C., Komsuoglu, H., Koditschek, D.E., Sensor data fusion for body state estimation in a hexapod robot with dynamical gaits, IEEE Trans. Robot., 22, pp. 932–943, 2006.

[9] Trifunovic, M., Vadiraj, A.M., Van Driel, W.D., MEMS accelerometers and their bio-applications, 13th International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE), 16–18 Apr. 2012, pp. 1–7, 2012.

[10] Stančin, S., Tomažič, S., Early Improper Motion Detection in Golf Swings Using Wearable Motion Sensors: The First Approach, Sensors, 12(6), pp. 7505–7521, 2013.

[11] Brutovsky, N., Novak, D., Low-cost motivated rehabilitation system for post-operation exercises, Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE, pp. 6663–6666, 2006.

[12] Zheng, H., Black, N.D., Harris, N.D., Position-sensing technologies for movement analysis in stroke rehabilitation, J. Med. Biol. Eng. Comp., 43(4), pp. 413–420, 2005.

[13] Liebermann, D.G., Berman, S., Weiss, P.L., Levin, M.F., Kinematics of Reaching Movements in a 2-D Virtual Environment in Adults With and Without Stroke, IEEE Trans. Neural. Syst. Rehabil. Eng., 20(6), pp. 778–787, 2012.

[14] Schall, G., Wagner, D., Reitmayr, G., Taichmann, E., Wieser, M., Schmalstieg, D., Hofmann-Wellenhof, B., Global Pose Estimation Using Multi-Sensor Fusion for Outdoor Augmented Reality. Proceedings of the 8th IEEE International Symposium on Mixed and Augmented Reality ISMAR 2009; Orlando, FL, USA. 19–22 October 2009, pp. 153–162. 2009.

[15] Su, C., Xu, W., Mengnan, G., Shengquan, Y., Simulation Teaching in 3D Augmented Reality Environment, 2012 IIAI International Conference on Advanced Applied Informatics (IIAIAAI), pp. 83–88, 2012.

[16] Stančin, S., Tomažič, S., Time- and Computation-Efficient Calibration of MEMS 3D Accelerometers and Gyroscopes,, Sensors, 14(8), pp. 14885–14915, 2014.

A Reasoning Geometric Modeling to Support Design for Dental Implant

Osiris Canciglieri Junior*, Anderson Luis Szejka*, Marcelo Rudek*, Teófilo Miguel de Souza**
* Pontifical Catholic University of Paraná / Polytechnic School - Production and System Engineering Graduate Program (PUCPR/PPGEPS), Curitiba, Paraná, Brazil.
** São Paulo State University – UNESP / Department of Electrical Engineering, Guaratinguetá, São Paulo, Brazil.
osiris.canciglieri@pucpr.br, anderson.szejka@pucpr.br, marcelo.rudek@pucpr.br, teofilo@feg.unesp.br

***Abstract* - The integration of different areas of knowledge to reach new technological solutions has become a reality. This integration has been improving the surgical process of dental implant applying the concepts and methods of the product engineering. In this context, this paper proposes a reasoning system that is capable to support a dental implant design for a single dental failure through CAD geometric modeling. This article has presented a case study of single dental failures that validate the reasoning system developed in the *Matlab* environment applying inside the Concurrent Engineering environment. The results demonstrated that the proposed reasoning system has potential to offer support to dentistry in the determination of the dental implant set more adequate to each patient.**

I. INTRODUCTION

The recent technological computer advances, in both hardware and software, has enabled a development in computer-aided design (CAD) from analysis to modeling. This evolution also allowed an opening for the integration of the engineering with other areas, particularly medicine and odontology in the bioengineering. According to [1], CAD systems have been extensively applied from prostheses and implants customization design until tissue engineering.

The medical imaging processing and concurrent engineering allied to computer aided design development have allowed an improvement in computer aided diagnosis [2], once information that is not obtained directly from the images can be extracted from image processing, such as density and bone geometry. The concurrent engineering systematizes the integration of these different tools in a tool to support the decision-making diagnostic processes and surgical procedures.

In dentistry the dental implant process is a multivariable process with a large dependence on the expertise of the dentist who will perform the procedure. Some computer systems help in the visualization of CT images obtained from patients, but do not provide essential information for planning the dental implant process and do not support the process of selecting the implants that best suits the patient causing premature failure and implants rejection. In the worst cases, during implant insertion it can interrupt a nerve and may result in partial or complete paralysis of the patient's mouth. Thus, this paper presents a reasoning system to support design for dental implant for determining implant that best suits the patient with a single missing tooth through the geometric modeling using CAD system, concurrent engineering environment and medical image processing. Also, the article presents the current state of the dental implant process, imaging recognition in computer tomography and its influences that were used as subsidies for the conceptual model development. The main contributions of this research can be highlighted using two cases study as: (i) Improvement of the dental implant process with decisions based in information extracted from image (patient's bone arch); (ii) reduction in the surgical time as well as in the implant absorption since there are fewer traumas; (iii) reduction in the risk of the dental implant rejection.

II. RESEARCH METHODOLOGY

This research is considered applied nature and qualitative approach as it searches to understand and explain knowledge for practical application oriented for the solution of specific problems through already existent theories and it seeks deep comprehension of a specific phenomenon through descriptions, comparisons and exploratory interpretations. The scientific aim of this research is exploratory and a literature review and experimental research were carried out. It is a literature review because based on the review of the literature it was built the knowledge to develop the methodological and experimental since it was necessary to determine the object of study and its variables making possible to control the object of study. The research's main objective is to propose a reasoning system using inference mechanism to structure logically the dental implant process to support the dentist's decision during the procedure. Thereby, for this purpose it is necessary to explore techniques of medical images processing in DICOM file format, protocols used in dental implant area. In the inference mechanism interaction, the model was validated by the implementation and tests in experimental case studies.

III. BIBLIOGRAPHIC BACKGROUND

The evolution of the computer systems is enabling the development of increasingly complex algorithms that perform processing almost instantaneously and with high degree of accuracy that are increasingly used in medicine and dentistry assisting in planning diagnostic and disease forecasting [3].

The computed tomography is a radiographic technique that consists in the acquisition of images in axial cuts that can be three-dimensional reconstructed [4], enabling

advances in imaging diagnostic, revolutionizing the practice of radiology, as well as the medicine and dentistry areas, combining techniques of image processing in the development of tools that provide medical data to assist in decision making processes [5].

DICOM standard made possible the image processing algorithms evolution since the information obtained from the hardware is the same, independently of the manufacturer. This fact allow the effort concentration on the development of systems to support doctors, dentists, nurses. Besides, DICOM image files can be converted in different formats, enabling the visualization in computers without dedicated application and the file image compression in order to send it to remote computers through the internet [6]. However, depending on the format choice there may be a considerable loss of important information for the image analysis [7].

In the oral implantology, dentistry branch for the edentulous treatment with the rehabilitation via dental implant it is verified an increment in the use of this equipment mainly in the 3D images reconstruction area through computed tomography, providing a better view to the dentist of the patient's bone structure. This overstepped some limitations in the conventional dental implant treatments planning mainly in pre-implantation stages, which used to be based on 2D data obtained by computed tomography. Thus, in this graphical multi-visualization environment proportionated by the image reconstruction there is an increase in the dentist interactivity with surgical planning, making the process increasingly safety and reliable [3].

In the implantology there is a division in the types of prostheses existing fixed and cemented prosthesis. The advantage of using the fixed prosthesis (screwed), according to [8] is the longevity that they present before the prosthesis partially fixed (screwed and cemented), since they reduce the risk of cavities; improve hygiene; reduce the risk of sensitivity and the contact with the root of the existing teeth; improve the aesthetics of the abutments; the cleaning of the bone in the edentulous space reducing the risk of prosthesis tooth loss besides the psychological aspect. As disadvantages are cited the high cost, high treatment time and the possibility of failures of implant insertion due to poor planning or execution. The advantage the non-occurrence of the process of reabsorption of the surrounding structures of the missing dental element, i.e. there is no absorption of the soft bone which is present in this region and thereby this research has chosen to use the fixed prosthesis [9].

The fixed prosthesis implanted can be divided into two main elements prosthesis segmented and non-segmented. Three distinct parts compose the segmented prosthesis: i) the implant; ii) the abutment; and iii) the crown. The non-segmented prosthesis consists of only two parts: i) the implant; and ii) the crown (built from a pillar connected to the prosthesis) facilitating the aesthetic result [10].

The use of the computed tomography in the dental implant process has made the procedure safer as in other areas that already use these images for the three-dimensional modeling (3D) as in the skull reconstruction where it is realized all the bone reconstruction and correct in virtual way all the missing parts that exists in the bone, exporting these information in CAD file, making possible the manufacture of the part and after that its insertion.

Another aspect to be considered is the use of simultaneous engineering concepts and computer aided diagnosis in the medical areas. The Institute Defense Analysis (IDA) defined the Simultaneous Engineering concepts as a systematic approach for the integration, product simultaneous conception and its related processes including the manufacture and support. This approach aims that the developers take in consideration, since the beginning, all the elements in the product life cycle [11]. In the medical sciences, the systems that use this philosophy improve the results as there is an investigation of different variables simultaneously that converges to a solution each time more reliable enhancing the diagnostics.

The integration of all these areas requires that the systems present expertise in the search of integration solutions as in the modeling problem solving attributing to the system. Thus, the conception of these systems, called specialized systems, can use the concept of inference mechanisms for structuring decision rules that helped in the solution of the multi-variable systems [12].

IV. Conceptual Structure of Design for Dental Implant System

The traditional dental implant procedure occur by visual analysis of CT or limited computational system but this process is not deterministic and no accurate make difficult and imprecise the dental implant definition which may cause its premature failure, bone loss, implant rejection and infection, as seen in the work of [12] and [13], compromising the treatment of partial and/or total edentulous patients. The dental implant procedure is multivariable and complex because all these variables should be analyzed simultaneously for design and determining the group of dental implant that is suitable.

There is wide range of data, which should be considered during dental implant design. These include bone information (density, type and geometry), nerves structure, dental implant data (type, model, diameter, length, density applied), but that traditional procedure is not analyzed with real data. However, this information can then be used by medical image processing and reasoning system to provide decision support to dentist surgeon, according to figure 1.

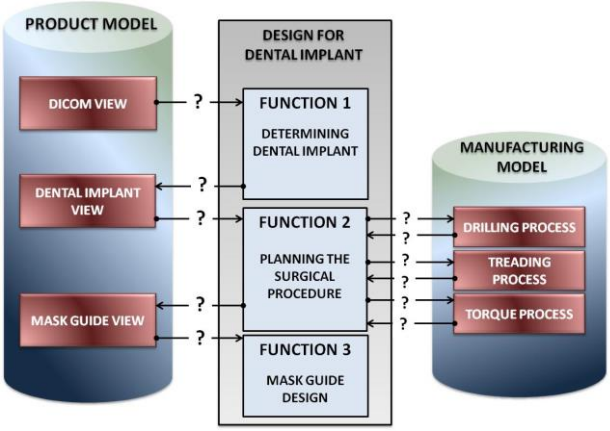


Figure 1. Conceptual Structure of Design for Dental Implant System

The reasoning system was proposed in the conceptual model using the concept of product model and design for

X. The first has information about the CT images and dental implant; and, the second, named Design Oriented for Dental Implant (DODI), has the product model for Dental Implant determination which contains the inference mechanisms that will convert, compare and share data.

The product model contains the informational requirements necessary to subsidize the reasoning system making processes of the DODI function such as tomographic images stored in DICOM representation and the information of the diameter, length, and density applied to the dental implant stored in the dental implant representation. The DICOM Representation contains information about the patient and CT image acquired from patients and stored in DICOM standard (figure 2). These images are gotten in cut axial, but to a complete analyses it is necessary other formats, thus there are image process where it is possible to get traverse cut. With these medical images are possible extract several information from processing image (geometry, density, nerves localization). The Dental Implant Representation – this representation presents the information about dental implants (model, type, diameter, length, density applied), which it can be to use to determine the group of implant that suits to patient and help the dentist surgeon during the surgery planning.

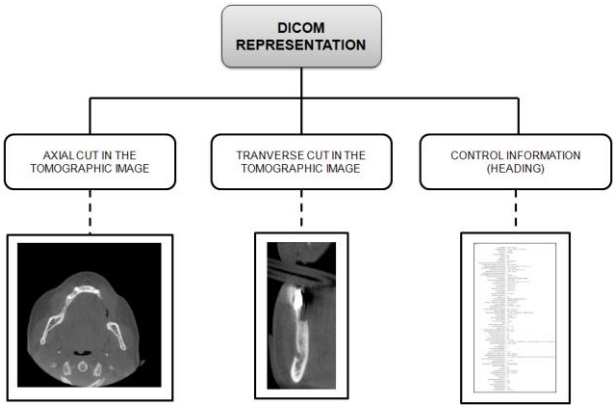


Figure 2. DICOM Representation Structure

Design Oriented for Dental Implant uses mechanisms to translate the know how of the surgeon dentist in structured rules (figure 3). The mechanisms are reasoning systems elements capable of searching the necessary rules for solving some problem. These rules are evaluated and ordain logically the heuristic process of inference allowing to the system searching these information and cross to a specific analyzing identifying the group of suits Dental Implant. The DODI was structured in function where each one has a group of inference mechanism responsible in providing solutions to determining the suits Dental Implant for the patient.

The function named "*Dental Implant Determination*" comprises the mechanisms that translate, convert and share the information contained in the DICOM representation (information of control; axial cut; transverse cut) in order to mathematically determine the parameters of diameter, length, bone density that will be used for the selection of the group of implants that best meet the desired requirements and that are available in the "*Dental Implants Representation*".

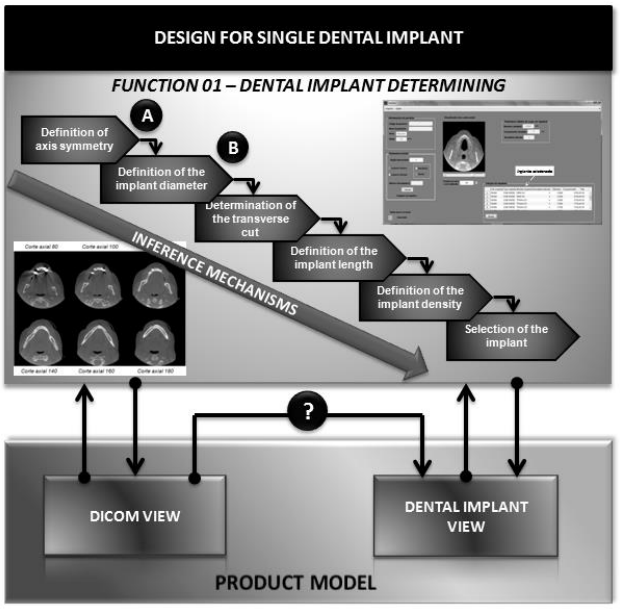


Figure 3. Design for Dental Implant System (single failure)

The definitions support the mechanisms to converting, translating and sharing information (data) between representations doing that the information of a representation offer support to decision in another representation. The conceptual model of DOID has inference mechanism allow interchange information as detail "?" and "→" in figure 3.

The mechanism "*Definition of Axis Symmetry*", selecting the region of interest and determining the geometric modeling of the symmetry axis, through treatment of the images and geometric analysis defined where occurs the insertion of the dental implant and allowing the conversion of axial cuts into transverse cut of this region. With this information, it is possible to identify the bone geometric and to define the limits of bone and teeth where this information is use to "*Definition of Implant Diameter*". The mechanism that allows change the image in axial cut to transverse cut along the symmetric line is "*Determination of transverse cut*" which is important to mechanism "*Define of implant length*" once it is possible analyses the bone depth and locating the nerves or fissure in the bone. Thus, with this details can determine the length of Dental Implant safety.

The mechanism "*Definition of Implant Density*" uses the information get for "*Definition of axis symmetry*" and "Determination of transverse cut" to realize an analysis for histogram to identify the concentration of similar levels pixels and determining what the type is the bone in that region. This is paramount importance because the first criterion to choice a Dental Implant has been known what density indicate.

The final mechanism "Selection of the Implant" is responsible to determine a Suitable Implant for a Single dental failure through the CAD Geometric Modeling. The dental implant process is multivariable and complex because multiple variables such as bone density, geometry of the dental arch, region of nerves, among others need to be analyzed simultaneously for determining the dental implant that is most suitable to the characteristics of the patient. The traditional dental implant procedures occur by

visual analysis of tomographic images or limited computational systems. According to [14] the existing systems do not provide sufficient informational subsidies for the correct determination of the dental implant, causing their premature failure. The non-accurate and reduced information make difficult and imprecise the dental implant definition which may cause its premature failure, bone loss, implant rejection and infection, as seen in the work of [13] and [12], compromising the treatment of partial and / or total edentulous patients.

The selection of the dental implant is a process of simultaneous and interdependent analysis of the aspects such as bone structure, nerves positioning, geometry of the mouth and teeth, and in the case of selection of single implants placed between teeth it is necessary to check the space available for inserting the implant with physical properties that can support the tooth masticatory. Figure 3 presents the conceptual methodological approach for determining the single dental implant, whose mark ("?") presents the necessary investigation to build an informational structure that provides support to the process of determining the dental implant.

This methodological proposal is divided in design system oriented to the process of single dental implant (DOSDI) and the product model where the latest provides informational support to the DOSDI, which uses the inference mechanisms concept for the determination of dental implants more adequate to the patient. Thus, these two elements, product model and DOSDI, is macro area that contains representations of various stages of the dental implant and their interdependence functions, containing all information relating to products, processes and procedures related to dental implant protocols.

The "*Design for Single Dental Implant*" mechanism was used for the development of DOSDI. The mechanisms are specialized systems elements capable of searching the necessary rules to evaluate and ordain logically the heuristic process of inference. This function, Dental Implant Determination - Function 01 (Figure 3), comprises the inference mechanisms that translate, convert and share the information contained in the DICOM Representation (information of control; axial cut; transverse cut) in order to mathematically determine the parameters of diameter, length, bone density that will be used for the selection of the group of implants that best meet the desired requirements and that are available in the Dental Implants Representation.

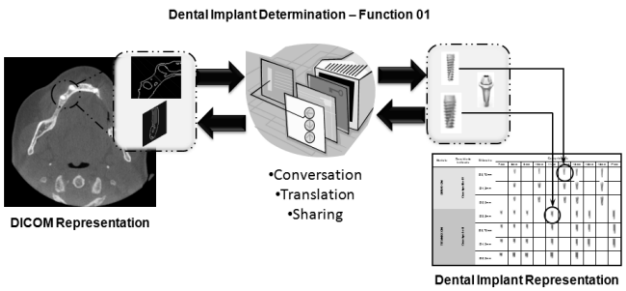


Figure 4. Dental Implant Determination Structure

The methodology used information conversion, translation and sharing mechanism contained in a representation to support the decision making function of determining the dental implant. Thus, the DOSDI was structured into six inference mechanisms: Definition of the region of interest and symmetry axis (Detail A - Figure 3); Definition of the dental implant diameter (Detail B - Figure 3); definition of the transverse cut; definition of the dental implant length; Definition of the bone density and selection of the dental implant. This article approaches the mechanisms for definition of the region of interest and geometric modeling of the symmetry axis and the mechanism of defining the implant diameter since these two mechanisms are responsible for the geometrical definition of the implant.

V. TRANSLATION MECHANISMS OF THE REGION OF INTEREST AND AXIS SYMMETRY

This mechanism is intended for selecting the region of interest and the geometric modeling of the symmetry axis (Figure 5), where occurs the insertion of the dental implant allowing the conversion axial cuts into transverse cut of this region. The definition of the region of interest is made by the oral facial dentist surgeon through the observation / analysis, identifying the image that presents in detail the gaps between two teeth. From this image the system intervenes and performs the geometric modeling of the dental arch.

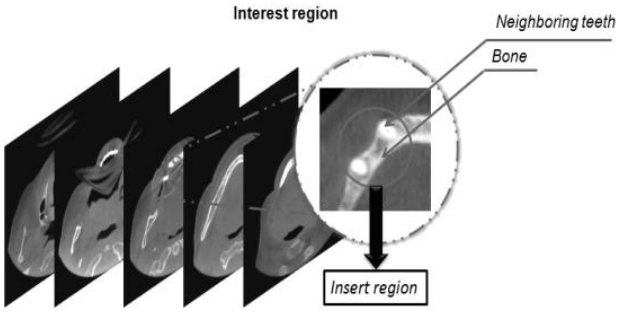


Figure 5. Region of interest selection

From the region of interest is extracted only the bone information for processing the images using as segregation parameter the Hounsfield scale. As a result of this process it is obtained only the information of the bone geometry and the failure surrounding teeth allowing a geometric analysis of the dental arch (Figure 6).

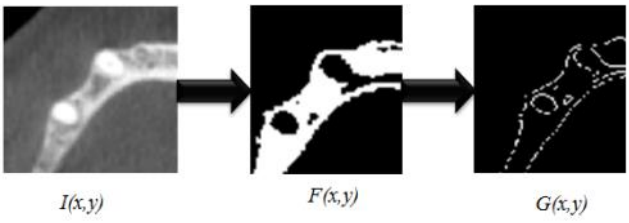


Figure 6. Detection of the bone geometry

In this analysis it is identified the geometrical center, through two reference lines that has are based on the inner edge the failure neighbors teeth and the intersection of these reference lines identifies the geometric center as shown in the Figure 7. Using this geometric center as a reference generates a symmetry line to the neighboring

teeth allowing the extraction of the insertion center and the implant diameter (Figure 8).

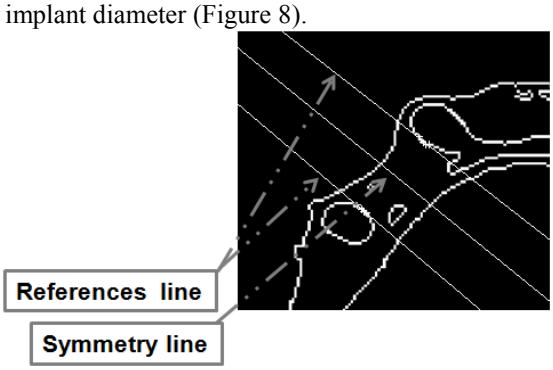


Figure 7. Symmetry axis construction

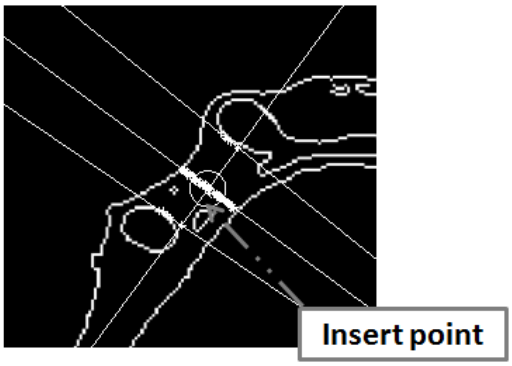


Figure 8. Identification of the insertion point

VI. Discussion of Results

The reasoning system development using experimental case allows the assessment of a partial edentulous with a single failure in the canine region (figure 9). The system outlined the axis of symmetry of the dental failure interacting with the dentist, creating an accurate axis, relying only on images with an uncertainty grade of 0.25 mm, which can be considered insignificant on the implantology optics. Beyond this point the reasoning system defined the implant diameter (TABLE I) based on the bone thickness obtained by geometric modeling, respecting the minimum area for the osseointegration that should be 1mm around the implant [15], [16] and [17].

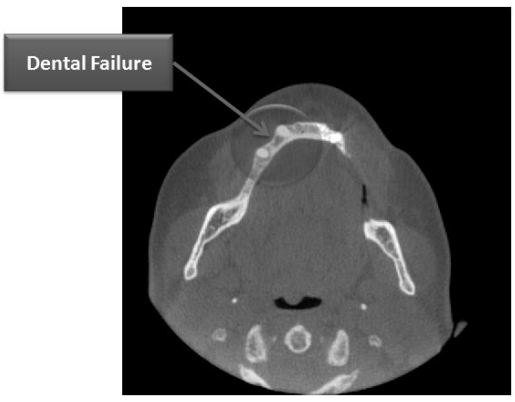


Figure 9. Dental failure between two teeth

For determining the implant diameter the system used the shortest distance between the outer edge of the bone and the internal bone, since this is smaller than the distance between the two teeth, obtaining as a result an implant of 3.85mm in diameter.

TABLE I.
IMPLANT MODELS OBTAINED FOR THIS DENTAL FAILURE
(FROM THE PROPOSED REASONING SYSTEM)

Type of Implant Body	Model of Implant Body	Density	Diameter of Implant Body (mm)
CONE MORSE	TITAMAXCM	2	3,5
CONE MORSE	TITAMAXCM	2	3,5
CONE MORSE	TITAMAXCM	2	3,5
CONE MORSE	TITAMAXCM	2	3,75
CONE MORSE	TITAMAXCM	2	3,75
CONE MORSE	TITAMAXCM	2	3,75
CONE MORSE	TITAMAXCM	2	4,0
CONE MORSE	TITAMAXCM	2	4,0
CONE MORSE	TITAMAXCM	2	4,0
INTERNAL HEXAGONAL	TITAMAXIIPLUS	2	3,75
INTERNAL HEXAGONAL	TITAMAXIIPLUS	2	3,75
INTERNAL HEXAGONAL	TITAMAXIIPLUS	2	3,75

With these data the system performs an enquiry to the dental implant representation database. It returns a selection of 12 models of dental implant that meet the design requirements, assigning to the dentist the identification of which implants are most likely to be used on the patient. In this case, knowing the manufacturer used in the research shows an availability of approximately 150 implant models, the system decreased by 92% the options, making the process more and more reliable.

VII. Conclusion

This paper presented a conceptual reasoning system that is able to determine the most suitable implant for single dental failure since the traditional procedures do not present informational requirements that support the dentist decision making process, leaving to him the selection of the best implant based on limited information. Thus, the proposal system was aimed to assist and support the process of determining the dental implant based on tomographic image processing and analysis of the existing models of dental implant, extracting the most important of characteristics them.

These support information have to be stored in the product model and the design function, via translation mechanisms, convert, share and translate the information from one representation to another, in order to select the set of implants best suited for the patient. As a result of this process, it was proposed an experimental case of single failure in the mandible at the canine region. It was obtained a set of 12 suitable implants to use in this particular case and it is up to the dental surgeon the identification of the more adequate implant for the patient.

The results showed the system's potentiality as a tool for computer aided diagnosis through the analysis of the bone geometry and other parameters making the dental implant procedure less traumatic, reducing the implant rejection level. For further researches, it is necessary: i) to explore in more detail the process of dental implants in total edentulous with the construction of a guide mask for the implant insertion procedure; and ii) to perform the correct planning of the dental implant inserting process identifying the depth of the insertion as well as the drilling and threading process.

ACKNOWLEDGMENT

The authors are thankful for the financial and technical support provided by Pontifical Catholic University of Paraná (PUCPR), Curitiba/Brazil.

REFERENCES

[1] W. Sun, B. Starly, J. Nam, A. Darling., Bio-CAD modeling and its application in computer-aided tissue engineering, Computer-Aided Design 37 (2005), 1097-1114.

[2] Z. Zhou, B.J. Liu, A.H. Le, CAD-PACS integration tool kit based on DICOM secondary capture, structured report and IHE workflow profiles, Computer Medical Imaging and Graphics 31 (2007), 346-352.

[3] D. Grauer, L. S. H. Cevidanes, W. R. Proffit, Working with DICOM craniofacial images. American journal of orthodontics and dentofacial orthopedics: official publication of the American Association of Orthodontists 136 (2009), 460-470.

[4] S. E. Duff, et al., Computed tomographic colonography (CTC) performance: one-year clinical follow-up. Clinical Radiology 61 (2006), 932-936. Jan. 2006.

[5] S. Jivraj, W. Chee, Rational for dental implants, British Dental Journal 200 (2006), 661-665.

[6] R. N. J. Graham, R. W. Perriss, A. F. Scarsbrook, DICOM demystified: A review of digital file formats and their use in radiological pratice. Clinical Radiology 60 (2005), 1133-1140.

[7] R. H. Wiggins, H. C. Davidson, H. R. Harnsberger , J. R. Lauman, P.A. Goede, Image file formats: Past, present, and future. RadioGraphics 21 (2001), 789-798.

[8] C.E. Misch, Implantes Dentários Contemporâneos. Santos Livraria. 2 nd ed. São Paulo-SP, Brasil, 2000.

[9] M. A. Bottino, M. K. Itinoche, L. Buso, R. Faria, Estética com implantes na região anterior. Revista Implant News 3 (2006), 560-568.

[10] K. Ochiai, S. Ozawa, A. A. Caputo, R. D. Nishimura, Photoelastic stress analysis of implant-tooth connected prostheses with segmented and non-segmented abutments. The Journal of Prosthetic Dentistry. 89 (2003), 495-502.

[11] D. Tang, L. Zheng, L. Zhizhong, L. Dongbo, S. Zhang, Re-engineering of the design process for concurrent engineering. Computer and Industrial Engineering. 38 (2000), 479-491.

[12] T. Li, et al., Optimum selection of the dental implant diameter and length in the posterior mandible with poor bone quality – A 3D finite element analysis. Applied Mathematical Modelling. 35 (2010), 446-456.

[13] A. D. Pye, D. E. A. Lockhart, M. P. Dawson, C. A. Murray, A. J. Smith, A review of dental implants and infection, Journal of Hospital infection 72 (2009), 104-110.

[14] C. G. Galanis, M. M. Sfantsikopoulos, P. T. Koidis, N. M. Kafantaris, P. G. Mpikos, Computer methods for automating preoperative dental implant planning: Implant positioning and size assignment. Computer Methods and Programs in Biomedicine 86 (2006), 30-38.

[15] Neodent, Catálogo de produtos (press) **1** (2011), 1-164.

[16] J. H. Lee, V. Frias, K. W. Lee, R. F. Wright, Effect of implant size and shape on implant success rates: A literature review. *Journal of Prosthetic Dentistry* **94** (2005), 377-381.

[17] J. Brink, S. J. Meraw, D. P. Sarment, Influence of implant diameter on surrounding bone. *Clinical Oral Implantology Res.* **18** (2007), 563-568.

Diagnosis of Lumbar Disc Herniation using Multilayer Perceptron Neural Network

Ivan Milanković [1, 2], Vesna Ranković [1], Miodrag Peulić [3, 4], Nenad Filipović [1], Aleksandar Peulić [1]
[1] Faculty of Engineering, University of Kragujevac, Serbia
[2] Research and Development Center for Bioengineering, BioIRC, Kragujevac, Serbia
[3] Faculty of Medical Science, University of Kragujevac, Serbia
[4] Clinical Center Kragujevac , Department for Neurosurgery, Kragujevac, Serbia
ivan.milankovic@kg.ac.rs, vesnar@kg.ac.rs, miodrag.peulic@gmail.com, fica@kg.ac.rs, aleksandar.peulic@kg.ac.rs

***Abstract*—The aim of this study was to develop multilayer perceptron (MLP) neural network model to predict lumbar disc herniation. The age, gender, body mass index, the maximum displacement of the body center of force, left foot center of force, right foot center of force in the x and y directions have been used as the input variables for the established MLP model. The measurements were performed using the commercial software Foot Work Pro. A total of 40 patients have been divided into training and testing data sets. The study results suggested that MLP would be an efficient soft computing tool for diagnosis of lumbar disc herniation. The Pearson coefficients have been computed as 0.941 and 0.938 for training and test data sets, respectively.**

I. INTRODUCTION

Medical decision-support systems are computer systems designed to assist physicians or other healthcare professionals in making clinical decisions [1]. The decision making process is a complex mechanism that has to take under consideration a variety of interrelated factors and functions [2].

In the recent decade, soft computing techniques are widely utilized to simplify the complex uncertainties which are present in the medical data as well as in the medical decision support systems. The advantages of using such intelligent systems include increasing speed of the diagnostic process saving clinician and patient time, and, at the same time, reducing time and improving the accuracy of diagnosis.

Recent years have witnessed the development of bioinformatics and medical informatics by using computational techniques for interpretation and analysis of medical data. Yardimci [3] has described a number of soft computing methods which incorporates neural networks, evolutionary computation, and fuzzy systems and theirs applications in medicine.

There are a wide variety of techniques for medical data classification. Simple classification algorithms include the distance-based classifiers (minimum distance, nearest neighbour) and the Beyesian classifiers, while more sophisticated approaches are based on support vector machines and neural networks.

In the study of Dreiseitl and Ohno-Machado [4], logistic regression and artificial neural networks models have been compared with other machine learning algorithms in medical data classification tasks. The results and performances of these models have been summarized.

Seera and Lim [5] have developed a hybrid intelligent system that consists of the fuzzy min–max neural network, the classification and regression tree, and the random forest model, as effective decision support tool for medical data classification.

Degenerative disc disease is the major abnormality that causes lumbar disc herniation. The most common current clinically approved standard for diagnosis is the magnetic resonance imaging procedure [6].

A Bayesian-based classifier with a Gibbs distribution for diagnosing lumbar disc herniation have been designed and implemented by Alomari et al. [7]. In this study 35 cases have been used for testing and 30 data cases have been used for training. The finally an average accuracy about 92.5% has been observed.

Shamim et al. [8] have used Fuzzy Logic-based fuzzy inference system for identifying patients unlikely to improve after disk surgery and explored FIS as a tool for surgical outcome prediction.

The main objective of this paper is to investigate the accuracy of multilayer perceptron neural network for diagnosis of lumbar disc herniation. The soft computing methodology-MLP model has been designed based on experimental data.

II. CLINICAL DIAGNOSIS OF LUMBAR DISC HERNIATION

The first, initial, symptom of lumbar disc herniation is usually the back pain which can be acute or chronic. In the most cases, the patients talk about localized back pain which spontaneously stops. The back pain can be very irritating. It can last for several weeks, after which it can evolve into debilitating pain which can irradiate into the legs. These symptoms can be followed with paresthesia and stiffness in the affected dermatome, as well as with muscular weaknesses [9].

Sometimes, the clinical diagnosis can be in the form of serious pains in the legs or even cramps, which come very soon after the initial symptoms. In other cases, the pain becomes intensified during the seating, standing, walking or even coughing and sneezing and any other sort of straining. The pain stops in the lying position with flexed hips and knees. The pains in the back and legs can persist at the same time, but in the most cases the back pain reduces with the appearance of sciatica [10,11]. It is likely that this phenomenon occurs due to the reduction of stretch fibers for pain in the annulus and the last ligament,

which occurs with a disc extrusion. Similarly, in rare occasions, serious sciatica symptoms may suddenly withdraw, but this phenomenon is usually associated with motor weaknesses and disorder sensibilities for the interruption function seriously compressed nerve root. In elderly patients, the pain in the legs often dominates over the back pain, from the very beginning of the disease.

III. FOOT PRESSURE MEASURING

AMCUBE Foot Work Pro, presented on Fig. 1, is a platform for the detection of foot pressure distribution in both, the dynamic, as well as static mode [12]. The platform is composed of 2700 capacitive pressure sensors that measure the pressure in the range from 10 KPa to 1200 Kpa, with an error less than 5% of the full range. The dimension of a sensor amounts to 7.6 mm x 7.6 mm, which represents the spatial resolution of the measurement. The sampling frequency is 100 Hz.

Figure 1. Platform for foot pressure distribution measuring - Foot Work

The static measurements consist in the fact that the subject stands on the surface for 10 seconds. During this period the Foot Work provides information about the foot pressure distribution which is presented on Fig. 2. It also provides information about the value of displacement of the center of mass of the whole body, the left and the right foot in the frontal and sagittal plane, which is presented on Fig 3.

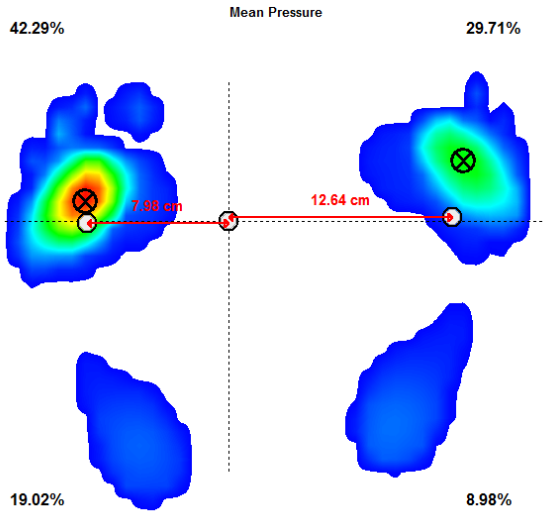


Figure 2. Distribution of foot pressure during the standing

The measurements were performed on a group of 40 subjects, aged between 18 and 73 years using the commercial software Foot Work Pro. We have presented a large number of measurements which included

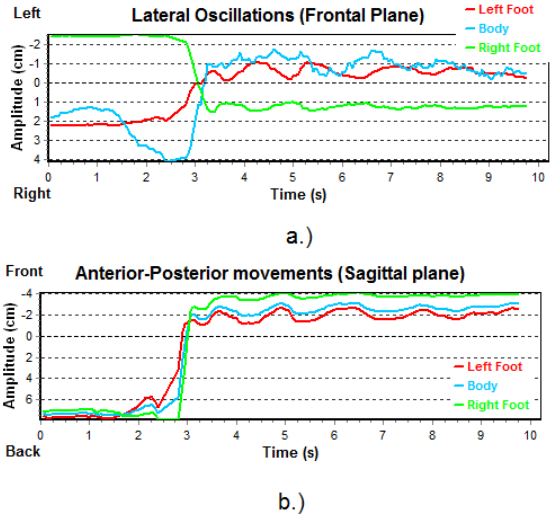


Figure 3. The value of displacement of the center of mass of the whole body, the left and the right foot in the a.) frontal plane and b.) sagittal plane

measurement of mobility during standing. Fig. 4 illustrates the domain displacement of the body center of force, left foot center of force, right foot center of force during the experiment. This domain is essentially a rectangle with sides representing the maximum displacements along the frontal and sagittal planes, respectively

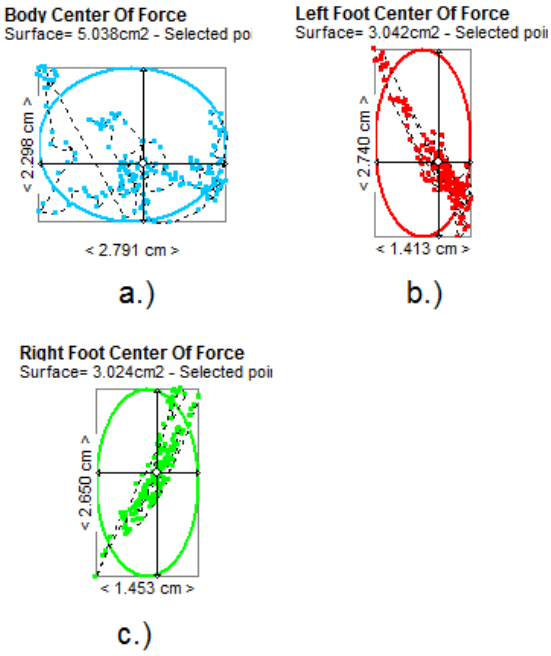


Figure 4. The spatial domain displacement of a)the body center of force, b)left foot center of force, c)right foot center of force

IV. THE ANN MODEL FOR DIAGNOSIS OF LUMBAR DISC HERNIATION

A. MLP neural network

A neural network consists of basic processing elements called neurons which are interconnected and distributed in layers. The neurons are represented by circles and the solid lines connecting the neurons represent weighting factors (Fig.5.).

The process to build MLP neural network model includes creating data sets for training and testing, training multiple MLP networks with varied parameters, and testing the models.

In this study, three-layered multilayer perceptron feedforward neural network was used and trained by the Levenberg–Marquardt algorithm. The nonlinear hyperbolic tangent sigmoid transfer function (tansig) and logarithmic sigmoid transfer function (logsig) were used in the hidden layer and the neuron outputs at the output layer. The hyperbolic tangent sigmoid transfer function and logarithmic sigmoid transfer function are described with the following equations:

$$\text{tansig}(u) = \frac{e^{u} - e^{-u}}{e^{u} + e^{-u}}. \quad (1)$$

$$\text{logsig}(u) = \frac{1}{e^{u} + e^{-u}}. \quad (2)$$

MLP is known to be universal approximators of the nonlinear functions, with a high degree of accuracy.

There are different algorithms for training MLP but the most often used is backpropagation rule However, backpropagation network has disadvantages of local minimum and slow convergence speed. The Levenberg–Marquardt algorithm is similar to the quasi-Newton method and one iteration of this algorithm can be written as:

$$\boldsymbol{\omega}_{k+1} = \boldsymbol{\omega}_{k} - \left(\boldsymbol{J}^{T}\boldsymbol{J} + \mu\boldsymbol{I}\right)^{-1} \cdot \boldsymbol{J}^{T}\boldsymbol{e}. \quad (3)$$

where $\boldsymbol{J}$ is the Jacobian matrix which contains first derivatives of the network errors with respect to the weights and biases, I is the identity matrix, μ is an adaptive factor and $\boldsymbol{e}$ is a vector of network errors.

B. Data set for lumbar disc herniation

In the MLP model for diagnosing lumbar disc herniation, one input, one hidden layer and one output layer are used as seen in Fig. 5. There are 9 inputs and 1 output for diagnosing lumbar disc herniation. Input variables names and their units are shown in Table 1. The outputs in the MLP model are LDH. The dataset has 40 observations used as training data. The first 30 data are from patients with lumbar disc herniation and others belong to people without lumbar disc herniation.

TABLE I.
INPUT VARIABLES NAMES AND THEIR UNITS.

Variable	Unit
Age	Numeric
Gender	Boolean
Body mass index	Weight (kg)'height (m)
Maximum displacement of the body center of force in the x directions	cm
Maximum displacement of the body center of force in the y directions	cm
Maximum displacement of the left foot center of force in the x directions	cm
Maximum displacement of the left foot center of force in the y directions	cm
Maximum displacement of the right foot center of force in the x directions	cm
Maximum displacement of the right foot center of force in the y directions	cm

C. Implementation of the MLP model

The optimal architecture of the MLP model for diagnosis of lumbar disc herniation was determined based on the maximum value of the Pearson coefficients of the training and test sets. The number neurons in the hidden layer was varied from 5 to 15 and the optimal number is chosen by trial and error approach, Table 2. Selection of an appropriate number of neurons in the hidden layer is very important aspect as a larger number of these may result in over-fitting, while a smaller number of neurons may not capture the information adequately.

The MLP developed in this research has 11 neurons in the hidden layer and one neuron in the output layer.

TABLE II.
THE PEARSON COEFfiCIENTS OF THE ARTIFICIAL NEURAL NETWORK MODELS WITH DIFFERENT NUMBER OF NEURONS IN THE HIDDEN LAYER

MLP–structure		Pearson coefficient
9 – 5–1	Training	0.923
	Test	0.914
9 – 7–1	Training	0.934
	Test	0.927
9 – 9–1	Training	0.938
	Test	0.929
9 – 11–1	Training	0.941
	Test	0.938
9 – 13–1	Training	0.925
	Test	0.921
9 – 15–1	Training	0.923
	Test	0.92

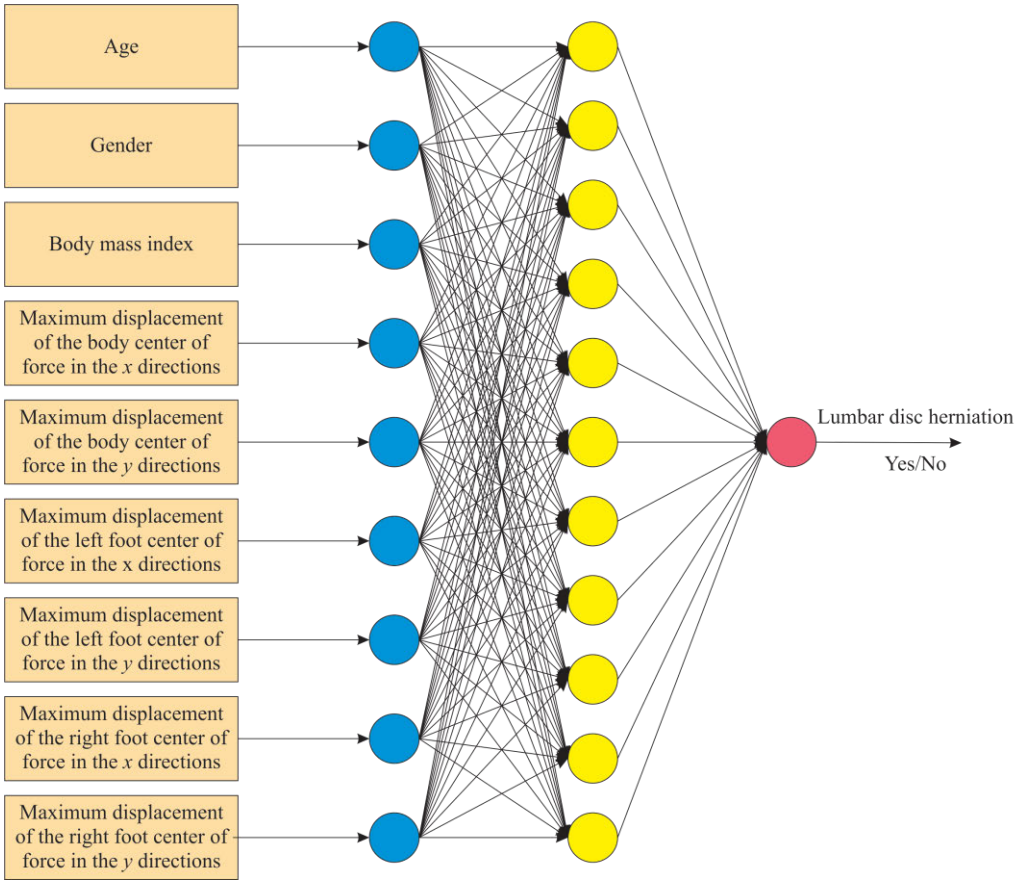


Figure 5. MLP model for diagnosing lumbar disc herniation

V. CONCLUSIONS

In this study, the applicability of the soft-computing methodology (MLP method) has been investigated for diagnosis of lumbar disc herniation. The performance of the neural network was measured using Pearson coefficient. The reported results confirmed that MLP model has good statistical performance.

The computational intelligence techniques can overcome the deficiencies of the conventional medical decision support systems that are based on statistical models. Medical decision support systems based on the soft computing technique can be seen to be a powerful alternative to traditional decision-making techniques.

Furthermore, for future research the prediction can be expanded with validation data on larger number of patients and to include other types of hybrid computationally intelligent systems (neuro-fuzzy system, genetic algorithm–fuzzy logic method, etc).

ACKNOWLEDGMENT

The part of this research is supported by Ministry of Science in Serbia, Grant III41007 and ON174028.

REFERENCES

[1] R. A. Miller, "Medical diagnostic decision support systems – Past, present, and future: A threaded bibliography and brief commentary," *Journal of American Medical Informatics Association*, vol. 1, pp. 8–27,1994.

[2] E.I. Papageorgiou, "A new methodology for Decisions in Medical Informatics using fuzzy cognitive maps based on fuzzy rule-extraction techniques," *Applied Soft Computing*, vol. 11, pp. 500–513, 2011.

[3] A. Yardimci, "Soft computing in medicine," *Applied Soft Computing*, vol. 9 , pp. 1029–1043, 2009.

[4] S. Dreiseitl, L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," *Journal of Biomedical Informatics*, vol. 35 , pp. 352 –359, 2002.

[5] M. Seera, C. P. Lim, "A hybrid intelligent system for medical data classification," *Expert Systems with Applications*, vol. 41, pp. 2239–2249, 2014.

[6] R. S. Alomari,, J. J. Corso, V. Chaudhary, G. Dhillon, "Toward a clinical lumbar CAD: herniation diagnosis," *International Journal of Computer Radiology and Surgery*, vol. 6, pp. 119–126, 2011.

[7] R. S. Alomari, J. J. Corso, V. Chaudhary, G. Dhillon, "Lumbar Spine Disc Herniation Diagnosis with a Joint Shape Model," *Lecture Notes in Computational Vision and Biomechanics*, vol. 17, pp 87-98, 2014.

[8] M. S. Shamim, S. A. Enam, U. Qidwai, "Fuzzy Logic in neurosurgery: predicting poor outcomes after lumbar disk surgery in 501 consecutive patients," *Surgical Neurology*, vol. 72, pp. 565 – 572, 2009.

[9] M. D. Charles Vega, "Spinal Surgery Superior to Exercise, Medical Therapy at 4 Years," *Medscape Medical News*; 2009.

[10] Ahn, Uri Michael, Ahn, Nicholas, Buchowski, Jacob M. Garrett, Elizabeth S. PhD; Sieber, Ann N. RN, MSN; Kostuik, John P. MD; "Cauda Equina Syndrome Secondary to Lumbar Disc Herniation: A Meta-Analysis of Surgical Outcomes;" *Spine,* vol. 25, no. 12, pp. 1515-1522, 2000.

[11] H.M.Mayer; "Minimally invasive spine surgery;" *Springer-Verlag Berlin Heidelberg;* 2006.

[12] www.amcube.net

Telerehabilitation Model of Physical Therapy using Kinect and Embedded Systems

S. Vukićević*
* School of Organizational Sciences, University of Belgrade, Serbia
vukicevic502011d@fon.bg.ac.rs

Abstract— **In this paper a model which provide the patient with a fast and simplified way of performing home therapy is described. Model is based on virtual reality, movement tracking and sensors' reading. Technically, it is consisted of (1) medically designed Software as a Service platform which provide remote, secure, reliable and always available software platform, (2) interactive virtual reality games that increase patient's motivation and concentration, (3) Microsoft Kinect for motion tracking and (4) embedded systems for tracking physical abilities during gameplay. The paper presents the results of applying this model of therapy to a single post-stroke patient. Therapy was focused on the upper limb and visual difficulty and resulted with improvements in both.**

Keywords – Motor Disorder, Telemedicine, Virtual therapy, Motivation, Post-Stroke

I. INTRODUCTION

According to World Health Organization Report [1], 10 million people survive a stroke worldwide each year, while half of survivors remain disabled. Stroke is one of the leading causes of disability. Paralysis of one side of the body is a frequent consequence, which is treated with physical therapy [2]. Recovery time is individual, but rehabilitation center accommodation time is limited and patients are forced to continue therapy at home. After leaving the hospital, recovery is reduced to occasional visits to rehabilitation center. Exercising at home depends on the patient's self-discipline. Repetition of same exercises leads to saturation and skipping therapy sessions, which usually leads to discontinuity of home physical therapy.

The goal of this paper is to create the system that would improve the success of treatment at home with small investments in equipment. Adding virtual reality to the system creates an interactive environment. Adding purpose to each movement, through interaction with virtual objects, the patient becomes less aware of physiotherapy. After the treatment, although physically exhausted, the patient will feel satisfaction for fulfilling tasks and will be more motivated to exercise again. Virtual physical therapy has the form of games because its several characteristics contribute to a better quality of exercise. In the games there is always the goal that player needs to reach, using the skills or capabilities, and there are rules and restrictions forcing player to perform actions properly.

This paper describes a complete model of virtual physical therapy for upper limb in gaming form, that besides maintaining the continuity of practice constantly records the results of exercises, generates reports and information about the success of therapy and automatically proposes new exercises if the patient shows good results in several consecutive therapy sessions. The model also has the ability to perceive limitations in hand movement and to adapt the exercises to the patient.

II. RELATED VIRTUAL REHABILITATION SOLUTIONS

With the advent of Microsoft Kinect, peripheral device for PC, in 2012, creating an interactive environment has become easier. Kinect can track movements of the whole body and therefore it is often applied in projects of virtual physical therapy. The Spanish group *VirtualWare* developed *VirtualRehab* system, which is consisted of control center as administrator software platform and several games designed for Kinect (http://www.virtualrehab.info). The control center is used by therapists to prepare a plan of exercises, monitor and assess the progress of therapy. Canadian startup *Jintronix* also developed rehabilitation software, based on Kinect that allows therapists to remotely monitor treatment and determine the following exercises based on the activities of the patient (www.jintronix.com).

Lithuanian start-up *Devmotion* goes a step further - they have virtualized entire area around person, which has been playing game. Their therapeutic solution was designed for children, in order to replace the virtual hospital ward (http://devmotion.eu/virtual-rehabilitation-solutions). It is not uncommon to develop specific wearable gadgets, for purpose of virtual therapy or specific measuring. Researchers at Britain's University of Southampton developed three tactile devices [3], which are stimulating finger skin, of patients with poorly-operated hand, in order to regain a sense of handling objects. Wearable, specially tailored gloves are also used in the treatment for hands and fingers. For instance, Swiss startup company *YouRehab* developed wearable interactive glove *YouGrabber,* for hand therapy as well as the *YouKicker* device, for leg therapy and foot movements [4]. Research [5], in which *CyberGlove* and *Rutgers Master II-ND* glove were worn by people with poor fingers mobility after a stroke, showed that the mobility of the thumb increased 50-140% and mobility of fingers increased 10-15%. However, the use of these wearable devices is not adequate for clinical conditions, due to sterilization and size that does not fit all patients, such as for children.

From the all above mentioned, Kinect is the basis of most virtual rehabilitation solutions. The exercises are usually performed through the game. In more serious projects, it is possible to remotely monitor the results of playing and adjust difficulty of the game. However, we

noticed that only few solutions accurately capture the mobility of limbs and improve capability and mobility of muscles. Also, feedback is usually calculated based on time spent playing game and on number of achieved goals, while the accurate measurements of muscle contraction or speed of a limb movements (reflex) is not considered. Projects' advantage are wearable embedded devices, which measure their physical ability and opportunities, while patient plays a game.

We believe that statistic data recording of each therapy, remote access by patient and therapist and remote control of therapy sessions by therapist, will contribute to better home therapy and continuity in exercises. Also, the patient could be aware of information doctor oversees and also follow games' results, which could contribute with confidence in telerehabilitation model.

III. TECHNICAL REQUIREMENTS OF THE MODEL

As a substitute for standard therapy, virtual therapy must ensure all benefits of standard therapy. The primary goal of virtual therapy is that patient performs the full day therapy. The secondary goal is to collect data during exercise, in order to measure progress, as well as to establish a relation between the type of exercise and mental, physical and verbal recovery. With these goals, a software-hardware model has been made (see Fig. 1), which is consisted of:

- *Kinect* motion tracking device,
- *Embedded system* for measuring physical parameters,
- *Software as a Service (SaaS)* cloud platform for performing therapy and processing, storing and analyzing data, collected during therapy sessions,
- *Interactive virtual reality games* that increase patient's motivation and concentration.

A. Motion Tracking

Kinect launched the expansion of various creative projects which use 3D camera. This device consists of RGB camera 640x480px resolution, infrared (IR) camera and IR projector. RGB camera is a standard color camera. IR projector emits infrared rays in space which bounce off objects and return back to IR camera. That's how the distance between Kinect and objects is measured. This feature is useful in the creation of therapy, if patient is required to put his hand in front or behind his body. All three components: RGB camera, IR camera and IR projector allow the creation of 3D images. Open source library, Simple-OpenNI, allows Windows, Linux and OSX users to connect with the Kinect device and thus opens the door to a wide range of artistic and medical projects. Kinect can distinguish parts of the body and determine the position and orientation of the body. In addition to objects, Kinect can detect sound. This feature is not sufficiently exploited, although Kinect can very accurately determine the source of sound.

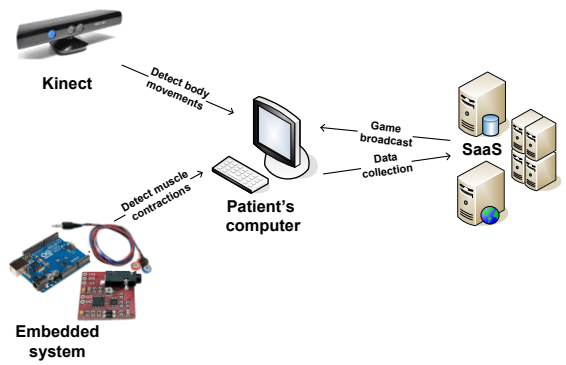


Figure 1. Telerehabilitation model of physical therapy

In our model, the function of Kinect is to determine the patient's mobility and limitations, before starting with virtual home therapy. By testing the patient on certain movements (e.g. height to which he can raise affected hand or move it to the right, left or how much he can bend it and how fast makes a move), a set of parameters is obtained, upon which the patient get a recommended list of games. During testing, body position is also very important, because patient will tilt to the right, if he finds too hard to lift left arm. Therefore, based on these initial measurements, a set of best suited games is automatically determined.

B. Embedded System forMeasuring Physical Parameters

Embedded system is a computer system of special purpose which executes predefined tasks. Embedded system is part of this model for two reasons: to diagnose the physical abilities of the patient based on measurements and to use values recorded from embedded devices, as input parameters in virtual games.

Embedded systems applied to this model are microcontroller Arduino Uno and muscle sensor with three electrodes. We can detect electrical potential EMG, using muscle sensor by placing electrodes in three positions: in the middle of the muscle, at the end of the muscle and on bony part near the muscle. Before placing electrodes it is necessary get skin prepared: remove hairs and clean it with alcohol. These steps are mandatory in order to provide better grip of electrodes and reduce the electrical resistance of the skin. Proper placement of EMG electrodes is very important for accurate measurement of muscle contraction. Unfortunately, if the muscle has more body fat, EMG signal will be weaker and difficult to record.

The value read from the muscular sensor will be used in rehabilitation games of this model. If the patient is required to alternately contracts and relaxes the muscle, it will represent an effort for him, but if he's doing the same unconsciously while playing game, the effect will be the same and results will be better. Based on the above, the use of muscle sensor in rehabilitation games should lead to improvements in patient's muscle structure.

Fig. 2 shows a successful connection of muscle sensor with Arduino Uno microcontroller. The sensor is supplied by two 9V batteries, because Arduino can provide a current of 40mA from digital pins, which is not enough. Arduino is connected to a computer using serial connection, but it is preferred to replace it with wireless connection. That is the way the patient would not be limited in space. Monitor displays muscle sensors measurements as we follow the contraction of the biceps. When the muscle is relaxed, third slider (blue) shows the

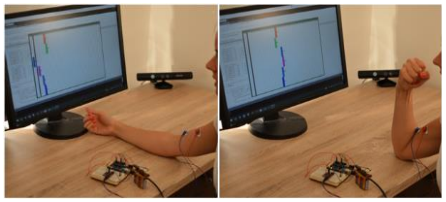


Figure 2. Example of reading data from the sensor muscle with electrodes connected to the microcontroller Arduino Uno. Relaxed biceps place blue slider to 0V (left), Contracted biceps place blue slider to 450V (right)

voltage 0V. When the biceps contracts, blue slider shows the voltage greater than zero. The voltage value depends on physical condition of the patient.

C. SaaS Cloud Rehabilitation Platform

"Cloud computing" refers to the use of the program and the storage space on a remote, scalable, virtual servers, which are considered safe and can be accessed at any time, from any location and any device. Users do not depend on the operating system; they are not required to install programs on personal computers, because programs are available via Internet. This virtual rehabilitation model contains a cloud distributed system, due to simplicity of communication between all users, including therapists, specialists, patients and their family. This SaaS rehabilitation platform consists of:

- *Web portal* for performing rehabilitation games and data access,
- *Data storage* for saving the results of sessions and storing individual patient's settings,
- *Web server* as a connection between data storage and web portal.

Data collected by Kinect and embedded system, during rehabilitation session, will be sent to the cloud SaaS platform. Web server will process received data, and keep every important piece of information in data storage. The problem may arise if user has slow Internet connection. Storing data at remote location (SaaS platform) can slow down game broadcast. The solution is to create local database on user's computer, which will be synchronized periodically with cloud storage.

Through web portal, patient or his family member must connect Kinect and embedded devices with SaaS web portal. Therapists can log into the same web portal and have the ability to view virtual record of patient, create treatment consisted of several games, modify games and communicate with patient.

Web portal contains the following modules:

- *Virtual record.* Each patient who uses SaaS rehabilitation platform will have a virtual record, which will keep information about every performed game including game settings, the goal to be achieved, time of completion, number of successfully and unsuccessfully completed tasks and muscle sensor measurements (muscle strength, reflex). Both, patient and therapist will be able to access all those data, at any time, from any device via Internet.
- *Automatic and manual game modification*. In addition to the manual modification by the therapist, game can be self-modifying. After testing the capabilities and limitations of patient's movements, game is automatically modified. For example, if patient needs to reach items above his head, but the patient is only able to raise his hand to the chin level, rehabilitation game is automatically adjusted to get items, which appear up to the chin of the patient. The therapist may recommend any rehabilitation game, if he determines that patient neglected certain types of exercises.
- *Online communication with therapist.* In addition to playing rehabilitation games, the model provides communication between patient and therapist via text messages, videos or online call.
- *Voice commands for patients with limited hand movements.* Taking into account that Kinect responds to voice commands, the patient can control rehabilitation session with his voice.

IV. Rehabilitation Games

The goal of the game in rehabilitation model is to decrease or eliminate the disability of the patient who survived a stroke, after which his psycho-physical abilities are reduced. Table 1 lists different types of difficulties and a set of movements that patient should practice through games. According to research [6] in United Kingdom, there are 77% post-stroke patients with upper limb disabilities and 72% of patients with lower limb disabilities, 60% of survivors have vision problems while 50% suffer slurred speech. These data indicate what type of rehabilitation games should be represented.

In addition to the games that promote physical and psychological condition of the patient, rehabilitation model includes games that measure progress in rehabilitation. The first group of games is performed every

TABLE I.
Difficulties and Movements Performed in Traditional Physical Therapy

Difficulty and % of affected people	Examples of motions rehabilitation games
Upper limb /arm weakness 77% [a]	Fetch the object in front of body, beside or above head, rowing, biceps contraction, swinging hands diagonally
Lower limb /leg weakness 72% [a]	Moving arm and leg to retrieve or avoid obstacles, bending at the knee, slight squats, lunge right-left, front-back
Visual problems 60% [a]	Tracking objects in particular color, drawing with eye movements, remove objects from the screen by looking at them
Slurred speech 50% [a], Reading difficulty	Reading text that is slowly slipping on the screen, reading the words that come out randomly on the screen
Trunk and postural control	Tilting to the right and left arm/elbow from a sitting position, rotation to the left or right from the sitting/standing position
Balance and reflexes	Exercises for static and dynamic balance, reaction to short-term events, getting up from a chair, lifting objects from the floor

[a] Based on Stroke association research "State of the Nation", United Kingdom, published in January 2015

day, while the second group is performed once a month or less frequently. This rehabilitation model is currently implemented in two games of the first group: "Tennis", intended for the treatment of arm weakness and game "Move items" for visual problems.

The game “Tennis” develops motor skills of stroke affected hand, especially the elbow and shoulder. Virtual tennis rackets are distributed on the screen. A tennis ball is moving toward one racket. The player should raise his hand to the height of the racket, at the time when ball comes to racket and clench the fist to catch the ball. The number of rackets is variable, from 2 for initial level of exercise to 9 for more advanced level. The game uses Kinect and embedded system described in chapter three. Kinect detects movements of the hand and the moment when the hand touches the racket. Embedded system can be placed on the biceps or forearm of affected arm. Embedded device is used only if the patient is able to contract the muscles of the hands. At the beginning of the game, patient is required to raise a hand and clench the fist, based on which the game automatically sets rackets at the proper height and determines whether it is possible to use the embedded system.

This virtual game with two rackets, left and right, was played by patient with very low mobility of the left hand. After initial measurements, left racket is set to the maximum height suitable for patient. Embedded device was not used because the patient was unable to clench the fist. The number of thrown balls per game was 60. The patient used his right hand to touch the racket containing ball on the right side of the screen and left hand to hit racket with ball on the left side of the screen. The results of playing are shown in Fig. 3, where it is noticeable that the number of errors decreases during 2 months of playing. The game can become more difficult if we speed up the balls or increase number of rackets.

The game “Move items” helps patients with reading problems caused by reduced field of vision. Patients with reduced field of vision in the left eye are not able to focus beginning of the line and they start reading from the mid-line. In order to improve the focus on the left or right visual field, a game called "Move items" is developed, where the user takes objects from one corner of the screen to another, see Fig. 4. Patients with reduced left field of vision will have a problem to focus objects on the left side of the screen, and they will need more time to transfer all balls from left to right side, rather than vice versa. Objects are moved by hand. Patient should use healthy hand.

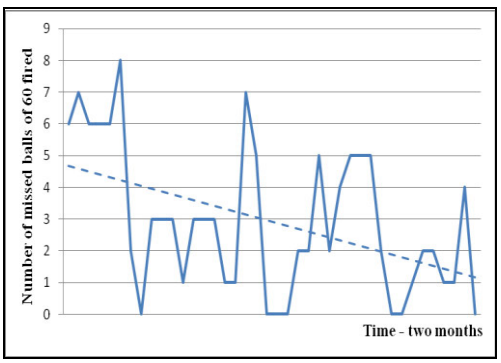
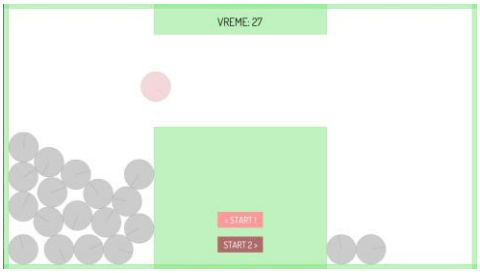


Figure 3: Progress in playing rehabilitation game “Tennis” by post-stroke patient with left arm weakness



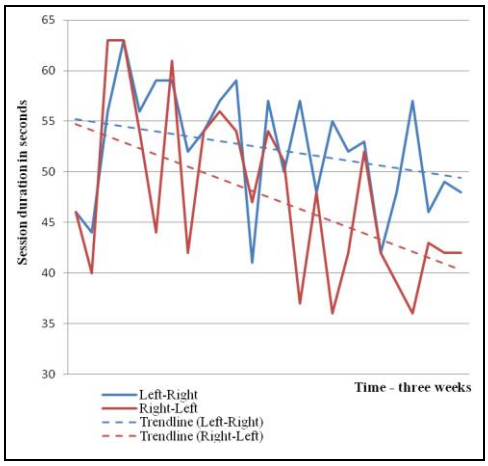


Figure 4: "Move items" game and progress in playing by patient with a reduced left field of vision. Red line - progress in focusing right side, blue line - progress in focusing the left side

Motion detection can be done by Kinect or touch screen. The game can be expanded with embedded sensor described in the third chapter, where fist clench can be used to lower the height of the barrier between left and right side, see Fig. 4, upper image.

Patient with reduced left field of vision played game “Move items” using right hand, in a three week period, once per day and the results are shown in the graph in Fig 4. Y axis shows the time spent on switching 20 balls from right to left (red line) and from left to right (blue line). The graph shows that transferring balls from left to right side lasts longer, which confirms that the patient slowly focuses beginning of the line (left side) during the reading. By comparing the measurements in the first five days and the last five days of the session, the duration of the session was reduced by 15% when moving twenty balls from right to left, and 27% when moving balls from left to right.

The other group of games is a substitute for therapist checking of the patient condition in rehabilitation center. The aim of this group of games is to determine current abilities of the patient and the progress of recovery after a period of practice. “Box and Block” test, published in 1957, is a method to evaluate the mobility of the upper limbs and coordination of movements. The box is placed in front of the patient. In the center of the box, barrier is positioned with cubes on one side of the box. The patient should take one cube at a time and pass over the barrier. If a cube drops out of the box on the other side of barrier, or if the patient moves two cubes over barrier, it is counted as one cube moved. The test is time limited, and progress in recovery is measured on basis of the number of transferred cubes. The game “Move items”, can be easily transformed into “Box and Block” test by modifying game parameters.

Fugl-Meyer test [7] is fundamental in measuring degree of impairment of motor function after stroke. Except for objectively measuring the damage, this test is used to periodically evaluate the degree of recovery of the patient. Reliability of Fugl-Mayer scale has been tested with success in many studies [8][9]. A large set of movements, performed during test, could be traced by Kinect.

V. CONCLUSION

The results of two pilot rehabilitation games showed noticeable improvements in the patient's home therapy. After weeks of playing, patient showed increase of concentration, faster reflexes, higher mobility of the affected hand and smoother reading.

The model described in this paper should enable faster recovery of patients who survived stroke. Kinect device is used as a sensor for the detection and tracking body segments. Embedded system records the muscle ability. Cloud architecture model provides remote access to rehabilitation system. This model eliminates the need for mandatory presence of therapist. The patient can perform all measurements at home, which reduces the cost of treatments. Therapist has access to patient's virtual record and may check its activities at any moment.

This model is developed for research and experiment purposes. The further development can create a product which will be widely used in post stroke telerehabilitation and evaluation of recovery degree.

REFERENCES

[1] The World Health Report 2002: "Reducing risk, promoting healthy life", http://www.who.int/whr/2002/en/.

[2] J. W. Sturm et al., "Handicap after stroke: how does it relate to disability, perception of recovery, and stroke subtype" in *Stroke*, 2002, vol. 33, pp. 762–768.

[3] G. V. Merrett, et al. "Design and qualitative evaluation of tactile devices for stroke rehabilitation", Assisted Living 2011, IET Seminar on. IET, 2011, pp. 1-6.

[4] K. Eng, K, et al. "Interactive Visuo-Motor Therapy System for Stroke Rehabilitation", in *Medical & biological engineering & computing* 2012, 45(9), pp. 901-907.

[5] R. Boian, et al. "Virtual reality-based post-stroke hand rehabilitation." Studies in health technology and informatics, 2002, pp. 64-70.

[6] http://www.stroke.org.uk/sites/default/files/State%20of%20the%20Nation_2015.pdf, Accessed: January 2015

[7] http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=908&Source=http%3A%2F%2Fwww%2Erehabmeasures%2Eorg%2FLists%2FRehabMeasures%2FAdmin%2Easpx. Accessed: January 2015

[8] P. W. Duncan, M. Propst, and S. G. Nelson, "Reliability of the Fugl-Meyer assessment of sensorimotor recovery following cerebrovascular accident," in *Physical therapy, 1983,* 63(10) pp. 1606-1610.

[9] D. J. Gladstone, C. J. Danells, and S. E. Black, "The Fugl-Meyer assessment of motor recovery after stroke: a critical review of its measurement properties", in Neurorehabilitation and Neural Repair, 2002, 16(3), pp. 232-240.

Prediction of wall shear stress in the arteries with myocardial bridge by neural networks

Dalibor Nikolić[2*], Igor Saveljić[1,2], Miloš Radović[1,2], Srđan Aleksandrić[3], Miloje Tomašević[3,4], Vesna Ranković[1], Nenad Filipović[1,2],

[1]Faculty of Engineering University of Kragujevac, Sestre Janjic 6, 34000 Kragujevac, Serbia
[2]Bioengineering Research and Development Center, Prvoslava Stojanovica 6, 34000 Kragujevac, Serbia
[3]Clinic of Cardiology, Clincal center of Serbia, Visegradska 26, 11000 Belgrade, Serbia
[4]Faculty of Medical Sciences, University of Kragujevac, Svetozara Markovića 69, 34000 Kragujevac, Serbia
markovac85@kg.ac.rs
isaveljic@kg.ac.rs
mradovic@kg.ac.rs
srdjanaleksandric@gmail.com
tomasevicmiloje@gmail.com
vesnar@kg.ac.rs
fica@kg.ac.rs

***Abstract*— Coronary arteries and their major branches, which supply oxygenated and nutrient filled blood to the heart muscle (myocardium), lie on the surface of the heart, in the subepicardial space, between visceral pericardium (epicardium) and myocardium. Sometimes, a shorter or longer segment of the epicardial coronary artery or its branch is covered by a band of heart muscle that lies on top of it. This band of muscle is called a "bridge" and the intramural segment of coronary artery a "tunneled artery".**

Myocardial bridging (MB) is a congenital coronary anomaly defined as a segment of a major epicardial coronary artery that runs intramurally through the myocardium beneath the muscle bridge.

It is very important to find the most efficient method for determining shear stress in the coronary arteries with myocardial bridge.

The procedure for calculating shear stress in MB arteries using neural networks trained with results from finite elements method will be explained in this paper.

I. INTRODUCTION

The value of shear stress in the artery with MB is very important for the medical doctors. Low and oscillatory shear stress, with a low time-averaged values (<1.5 N/m2), lead to alterations in the expression of vasoactive agents, such as endothelial nitric oxide synthase (eNOS), endothelin-1 (ET-1), angiotensin-converting enzyme (ACE) and growth-promoting and prothrombotic phenotype, ultimately acquiring a predisposition to atherosclerosis [1, 3, 4, 5]. On the other hand, the normal shear stress, with a positive time-average ranging between 1.5 N/m2 and 7.0 N/m2, increases the production of nitric oxide (NO) in endothelial cells and downregulates the expression of proatherogenic molecules, related to an atheroprotective effect [1,3,4,5]. In addition, scanning electron microscopy reveals the changes in the shape of the endothelial cells in LAD intima from flat and polygonal in the segment proximal to the MB to helical, spindle-shaped under the MB [2,6,7].

Since it is not possible to measure shear stress in the artery with myocardial bridge, it is very important to find the method for calculating this value.

One of most effective methods is FEM (Finite Element Method). Since it is a very complicated and slow process of mesh generating and solving which requests a very high computational power, we try to find some faster and easier methods to use, such as neural networks.

II. METHODS AND MATHERIALS

A. Developing geometrical FE model Artery withMB

FE model of artery with MB is created in our software for generating and meshing finite element models. After meshing, the application automatically runs the FE solver, waiting for the results and then imports them and creates a file for training neural networks. Application block diagram is presented in the Figure 1. [13].

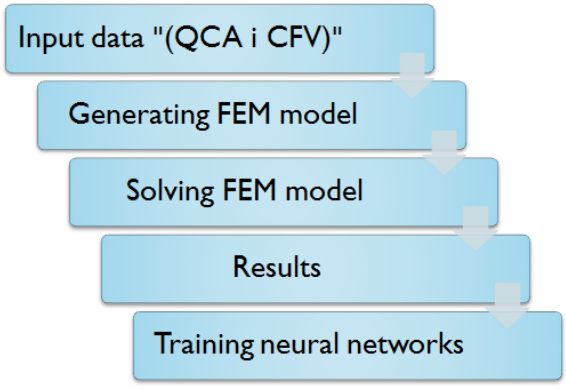


Figure 1. Block diagram

Segmentation of CT images is the best method to create the geometry of certain organ, but MB is very rare disease and in the medical practice, it is detected by angiography. For this reason, the software application is developed to generate the geometry from QCA measurements data from angiographic images.

Time interval - the end of systole			
Length MB	The diameter of the artery in the MB	The diameter of the artery in front of the MB	The diameter of the artery behind the MB
Time interval - early diastole			
Length MB	The diameter of the artery in the MB	The diameter of the artery in front of the MB	The diameter of the artery behind the MB
Time interval - mid-diastolic			
Length MB	The diameter of the artery in the MB	The diameter of the artery in front of the MB	The diameter of the artery behind the MB
Time interval - end-diastolic			
Length MB	The diameter of the artery in the MB	The diameter of the artery in front of the MB	The diameter of the artery behind the MB
Inlet velocity			

Figure 2. Input data from QCA

Additionally, the software automatically generates four geometrically different meshes for each of the measured periods of a heart cycle. (Figure 2). Based on these four meshes, the software interpolates the shape of the mesh model throughout the cardiac cycle. This mesh movement is very important for accurate computation of blood flow through MB.

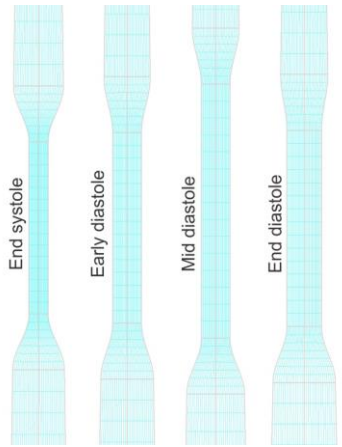


Figure 3. Generated FE 2D model meshes during 4 periods of a heart cycle

B. FEM mathematical method

A finite element model of the bridge was employed. Flow in the coronary artery is a complex, time-dependent, three-dimensional flow. The time-dependent and full three-dimensional Navier-Stokes equations have to be solved. A finite element mesh with 8247 2D 4 node axi-symmetric finite elements was generated using an automatic mesh generator as it is presented in the Figure 3. The mesh independence was reached at 6530 to 25375 finite elements.

The three-dimensional flow of a viscous incompressible fluid considered here is governed by the Navier- Stokes equations and continuity equation that can be written as

$$\rho\left(\frac{\partial v_i}{\partial t}+v_j\frac{\partial v_i}{\partial x_j}\right)=-\frac{\partial p}{\partial x_i}+\mu\left(\frac{\partial^2 v_i}{\partial x_j\partial x_j}+\frac{\partial^2 v_j}{\partial x_j\partial x_i}\right) \quad (1)$$

$$\frac{\partial v_i}{\partial x_i}=0 \quad (2)$$

where v_i is the blood velocity in direction xi, ρ is the fluid density, p is pressure, μ is the dynamic viscosity; and summation is assumed on the repeated (dummy) indices, i,j=1,2,3. The first equation represents balance of linear momentum, while the equation (2) expresses incompressibility condition.

Each waveform of and pulsatile flow was discretized into 500 uniformly spaced time steps. In the analysis, it was considered that the convergence was reached when the maximum absolute change in the nondimensional velocity between the respective times in two adjacent cycles was less than 10^{-3}.

The code was validated using the analytical solution for shear stress and the velocities through curve tube [8]. The pressure is eliminated at the element level through the static condensation.

A standard Petrov-Galerkin upwind stabilization technique was used for Re number [8].

In addition to the velocity field, the wall shear stress computation was performed. The mean shear stress τ_{mean} within a time interval T is calculated as [9]

$$\tau_{mean}=\left|\frac{1}{T}\int_0^T t_s\,dt\right| \quad (3)$$

where t_s is the surface traction vector. Another scalar quantity is a time-averaged magnitude of the surface traction vector, calculated as

$$\tau_{mag}=\frac{1}{T}\int_0^T |t_s|\,dt \quad (4)$$

Also, a very important scalar in the quantification of unsteady blood flow is the oscillatory shear index (OSI) defined as [9]

$$OSI=\frac{1}{2}\left(1-\frac{\tau_{mean}}{\tau_{mag}}\right) \quad (5)$$

$$\mathrm{Re}sT=((1-2\cdot OSI)\cdot\tau_{mag})^{-1} \quad (6)$$

In order to make mesh moving algorithm we implemented ALE (Arbitrary Lagrangian Eulerian) formulation for fluid dynamics [10]. The governing equations, which include the Navier-Stokes equations of balance of linear momentum and the continuity equation, can be written in the ALE formulation as [10].

$$\rho\left[v_i^*+\left(v_j-v_j^m\right)v_{i,j}\right]=-p_{,i}+\mu v_{i,jj}+f_i^B \quad (7)$$

$$v_{i,i}=0 \quad (8)$$

where v_i and v_i^m are the velocity components of a generic fluid particle and of the point on the moving mesh occupied by the fluid particle, respectively; ρ is fluid density, p is fluid pressure, μ is dynamic viscosity, and f_i^B are the body force components. The symbol "*" denotes the mesh-referential time derivative, i.e. the time derivative at a considered point on the mesh,

$$(\,)^* = \frac{\partial(\,)}{\partial t}\Big|_{\xi_i=const} \tag{9}$$

and the symbol "$_{,i}$" denotes partial derivative, i.e.

$$(\,)_{,i} = \frac{\partial(\,)}{\partial x_i} \tag{10}$$

We use x_i and ξ_i as Cartesian coordinates of a generic particle in space and of the corresponding point on the mesh, respectively. The repeated index means summation, from 1 to 3, i.e. j=1,2,3 in Eq. (7), and i=1,2,3 in Eq. (8). In deriving Eq. (7) we used the following expression for the material derivative (corresponding to a fixed material point) $D(\rho v_i)/Dt$,

$$\frac{D(\rho v_i)}{Dt} = \frac{\partial(\rho v_i)}{\partial t}\Big|_{\xi} + (v_j - v_j^m)\frac{\partial(\rho v_j)}{\partial x_i} \tag{11}$$

The derivatives on the right-hand side correspond to a generic point on the mesh, with the mesh-referential derivative and the convective term.

C. Neural networks

The results from FEM models are used for training neural networks. For this research we made calculation for 12 different patients with MB. In order to train the neural networks, a large number of patients were required.
For these purposes, input data for 2000 patients were randomly generated, based on the data from real patients with deviation 30%.

D. Neural network model

A graphical structure of neural network used for the prediction of the shear stress on arteries with MB is presented in the Figure 4.

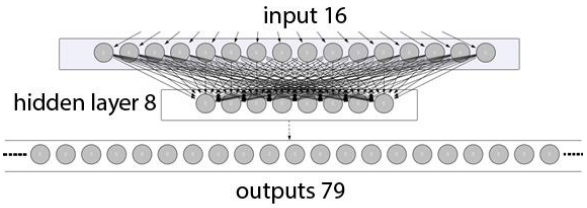


Figure 4. Graphical representation of neural network

E. Software solution for neural networks

To generate and training (Figure 5) neural network is used a software package MATLAB v.13

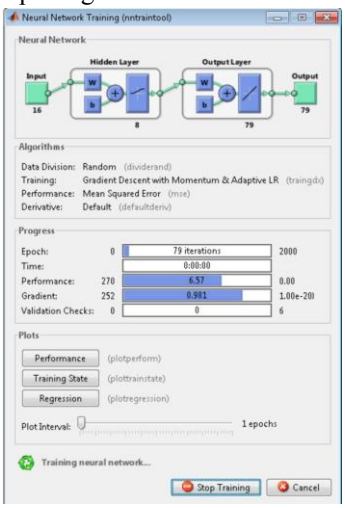


Figure 5. MATLAB Neural network training

F. Methodology neural networks

The traditional form of the back propagation algorithm has a problem with the local minimums and slow convergence. In order to overcome these problems, numerous variations of this algorithm have been developed [14]. For prediction of WSS, REST and OSI the multilayer perceptron neural network was used. This network was trained with back propagation algorithm. The adaptation of weight coefficient

$$w_{i,j(l)}(t+1) = w_{i,j(l)}(t) + \Delta w_{i,j(l)}(t+1)$$
$$w_{i,j(l)}(t+1) = \eta\alpha\frac{\partial E_k}{\partial w_{i,j(l)}} + \alpha\Delta w_{i,j(l)}(t) \tag{12}$$

The learning speed η in the equation (12) is variable. In every epoch, if the criteria function is decreased towards the aim, the learning speed is increased for factor η_{inc}. The values of the parameters α, η_{inc}, η_{dec} and max_{inc} used for solving the problem are presented in Table I.

TABLE I.
BACK PROPAGATION ALGORITHM PARAMETER VALUES

α	η_{inc}	η_{dec}	max_{inc}
0.9	1.05	0.7	1.04

The criterion for learning stopping is defined as 2000 epochs.

III. RESULTS

The neural network was tested by the *10-fold cross validation* algorithm and by calculating relative mean squared error - RMSE:

$$RMSE = \frac{\sum_{i=1}^{N}\sum_{j=1}^{S}(y_i(j)-y_i(j))^2}{\sum_{i=1}^{N}\sum_{j=1}^{S}(y_i(j)-\overline{y}(j))^2} \quad (15)$$

where N is the total number of the examples (2000), s is the total number of the outputs (79), $y_i(j)$ is the real value of the j-th output, i-th example, $\overline{y}(j)$ is the mean value of the j -th output, and $\hat{y}(j)$ is the neural network prediction of the j-th output, i-th example.

RMSE values obtained by neural network testing are presented in Table II.

TABLE II.
RMSE VALUES OBTAINED BY TESTING OF THE NEURAL NETWORK

Model	RMSE
WSS	0.0964
ResT	0.2873
OSI	0.1507

The value of RMSE which is lower than 1.0 shows that the model is usable (it has a lower error than non-intelligent model) [15]. From Table II, it can be concluded that the neural network has presented a high potential for the prediction of WSS, ResT, OSI.

IV. CONCLUSION

The advantage of neural networks lies in the fact that when one is trained, based on several input parameters, it provides very accurate results in just a few seconds, which is crucial in the clinical practice. However, the biggest disadvantage is that for training the network we must use the results previously obtained from the methods such as finite element method provided by numerous simulations.

This process of preparation of the results for neural networks training is very slow and it requires a high computing power to solve all of the FEM models and a trained person with the experience in FEM method.

V. ACKNOWLEDGMENT

This work was supported in part by grants from Serbian Ministry of Education and Science III41007, ON174028.

REFERENCES

[1] JR. Alegria, J. Herrmann, DR Jr Holmes, A. Lerman, SR. Charanjit. Myocardial bridging. Eur Heart J, 2005, 26:1159–1168.

[2] S. Möhlenkamp, W. Hort, J. Ge, R. Erbel. Update on Myocardial Bridging. Circulation, 2002, 106;2616-2622

[3] YS. Chatzizisis, GD. Giannoglou, Myocardial bridges are free from atherosclerosis: Overview of the underlying mechanisms. Can J Cardiol, 2009, 25:219-222.

[4] AM. Malek, SL. Alper, S. Izumo, Hemodynamic shear stress and its role in atherosclerosis. JAMA 1999, 282:2035-42.

[5] YS. Chatzizisis, AU. Coskun, M. Jonas, ER. Edelman, CL. Feldman, PH. Stone, Role of endothelial shear stress in the natural history of coronary atherosclerosis and vascular remodeling: Molecular, cellular and vascular behavior. J Am Coll Cardiol, 2007, 49:2379-93.

[6] Y. Ishikawa, Y. Kawawa, E. Kohda, K. Shimada, T. Ishii. Significance of the anatomical properties of a myocardial bridge in coronary heart disease. Circ J, 2011, 75:1559-1566.

[7] T. Ishii, Y. Hosoda, T. Osaka, T. Imai, H. Shimada, A. Takami, H. Yamada. The significance of myocardial bridge upon atherosclerosis in the left anterior descending coronary artery. J Pathol, 1986, 148:279–291.

[8] N. Filipovic, M. Kojic, M. Ivanovic, B. Stojanovic, L. Otasevic, V. Rankovic, MedCFD, Specialized CFD software for simulation of blood flow through arteries, (University of Kragujevac, 34000 Kragujevac, Serbia, 2006)

[9] C.A. Taylor, T.J.R Hughes and C.K. Zarins, Finite element modeling of blood flow in arteries, Comput. Meths. Appl. Mech. Engrg 158, 1999, 155-196.

[10] N. Filipovic, S. Mijailovic, A. Tsuda, M. Kojic, An implicit algorithm within the Arbitrary Lagrangian-Eulerian formulation for solving incompressible fluid flow with large boundary motions, Comp. Meth. Appl. Mech. Eng 195, 2006, 6347-6361.

[11] J. Ge, R. Erbel, H.J. Rupprecht, L. Koch, P. Kearney, G. Gorge, M. Haude, J. Meyer, Comparison of intravascular ultrasound and angiography in the assessment of myocardial bridging. Circulation 89, 1994, 1725–1732.

[12] A.M. Malek, S.L. Alper, S. Izumo, Hemodynamic shear stress and its role in atherosclerosis, JAMA 282, 1999, 2035-2042.

[13] D. Nikolić, M. Radović, S. Aleksandrić, M. Tomašević, N. Filipović, Prediction of coronary plaque location on arteries having myocardial bridge, using finite element models, Computer Methods and Programs in Biomedicine, Vol. 117, Issue 2, 2014, pp. 137-144

[14] M.B. Hudson, M. T. Hagan, H. B. Demuth, Neural Network Toolbox User's Guide, The MathWorks, Inc, http://www.mathworks.com/help/pdf_doc/nnet/nnet_ug.pdf

[15] I. Kononenko, M. Kukar, Machine learning and data mining, Horwood Publishing Chichester, UK, 2007.

Designing of Internal Dynamic Tibia Fixation 3D Model according to Mitkovic type TPL

Miodrag Manić*, Milorad Mitković**, Zoran Stamenković*, Nikola Vitković*
*University of Niš, Faculty of Mechanical Engineering/Department for Production, IT and Management, Niš, Serbia
**University of Niš, Faculty of Medicine/Department for Surgery, Niš, Serbia
miodrag.manic@masfak.ni.ac.rs
mitkovic@gmail.com
zoki2101984@gmail.com
nvitko@gmail.com

***Abstract* — This paper presents a display of originally developed method for designing a 3D model of internal dynamic tibia fixation according to Mitkovic type TPL. The internal side of the fixation, the one lying on the bone, is fully aligned with the anatomical shape of the bone surface. The method is based on the application of parameter 3D model with the marked bone fractures. This method can also be used for any other bone or tile type implant. The given model is designed for fixation production utilizing any method, and it is ideal for 3D printing.**

I. INTRODUCTION

In orthopedic surgery it is of paramount importance to apply proper methods of human skeletal system fixation in order to treat various bone fractures or other traumas. In case of internal fractures fixation treatment, it is very important to utilize internal fixation whose geometrical and topological characteristics fully correspond to the shape and size of the patient's bone, since this ensures faster and better recovery.

In order to reach this goal, the method which allows 3D modeling of internal fixation according to Mitkovic type TPL (tibia-plato-lateral) was created [1]. The suggested method is original and is based on designing fixation contour directly on patient's bone, after which extrusion is performed in order to create 3D fixation model. When this is finished, further model adjustments to the patient's bone are performed.

This approach enabled surgeons to create the fixation surface lying on the bone which fully corresponds to anatomical shape of the bone surface. The 3D model of the patient's bone which is needed for this procedure can be produced with any given method [2].

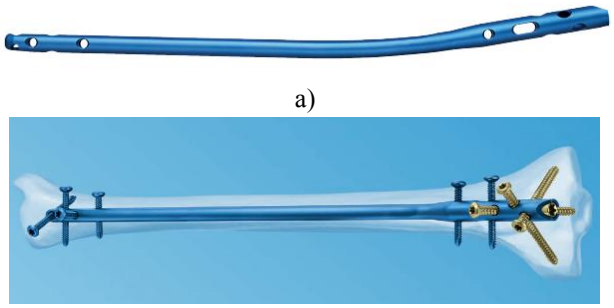
When fixation model is created in this way, it can be used for fixation production on 3D printer or CNC machines.

The method developed in this paper can also be used for various kinds of internal fixations that are directly attached to any bone surface.

II. SHORT DESCRIPTION OF INTERNAL FIXATION, SHAPES AND DESIGNING METHODS

Internal fixations are medical devices used as support to treat damaged or disease-infected bones brought about as a consequence of old age, disease or an accident. They are made of different kinds of biocompatible materials [3]. There are two kinds of internal fixations – intramedullary and extramedullary.

Intramedullary internal fixations are actually implant nails, as seen in Figure 1a, that are used to treat various bones (eg. tibia). They are inserted into the bone by using the bone's intramedullar canal, after which the screw bolts are inserted through the bone and previously made transverse screw holes on a nail. In this way, the binding of fractured bone fragments into a whole is created in order to treat the fracture and enable the bone to heal properly (Figure 1b) [4].



a)

b)

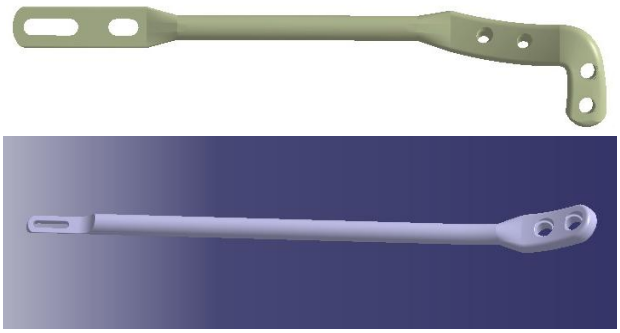
Figure 1. a) Intramedullary nail for tibia; b) Nail inserted into a bone [4]

Extramedullary fixations is comprised of various implants: screw bolts, tiles and dynamic fixations according to Mitkovic of different shapes and dimensions (Figure 2). These kind of implants are placed on the external surface of the fractured part of the bone [3[[1].

After this process is finished, the system of screw bolts is inserted through the previously made screw holes on a fixation. In this way, fractured bone fragments are connected into a whole, transport capacity of the joint is created and position and direction of the fragments are kept (Figure 3).

What is of paramount importance during the process of implants insertion is to create minimal direct contact between the fixation and the bone surface, while at the same time ensure that fixation follows bone contour. The pressure created when fixation rests on the bone should be avoided since it can damage the periosteum which covers the bone surface and nourishes the bone through blood vessels in it.

a)



b)

Figure 2. a) Standard and locking tiles and screw bolts.; b)Tibia fixations according to Mitkovic

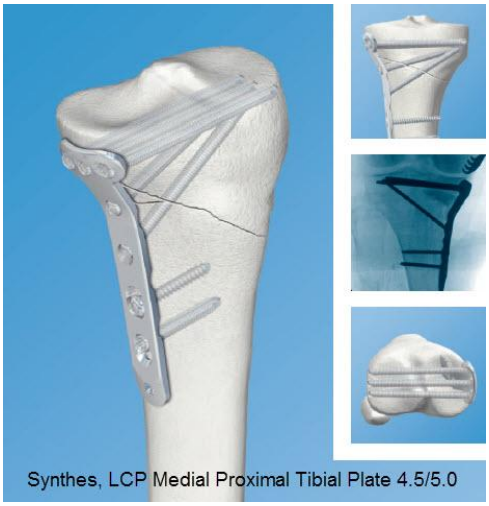


Figure 3. Insertion of a tile for the upper part of tibia [4]

III. 3D MODELING OF INTERNAL IMPLANTS

There can be found only few described examples of the methods for creating a 3D model of anatomically adjusted internal implants which correspond to the bone contour in the literature.

Usual methods that are applied include the using of CAD software to plan and design internal implants and these methods are based upon the ideas of orthopedic surgeons and engineers. These people constantly seek for new methods and techniques for designing and production of the internal fixations that could heal any bone fracture.

The basis for creating a 3D geometrical model of an internal fixation is a scheme of its contour defined in a suitable position in relation to the bone surface. Using the scheme, retraction or rotation of the model volume is performed, depending on the type of the internal fixation that is being used. Furthermore, modeling of the contact surface of the fixation, the surface lying on the bone, is performed In the end, new screw holes for the corresponding screw bolt type are made on the surface of the fixation model.

In [5] it can be seen an example of designing a fixation in the shape of a tile (Figure 4), as well as the dynamic screw bolt of a hip [6] (Figure 5).

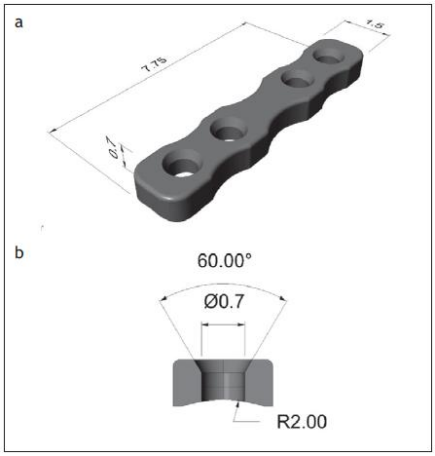


Figure 4. a) Tile modeling; b) Creation of the screw holes [5]

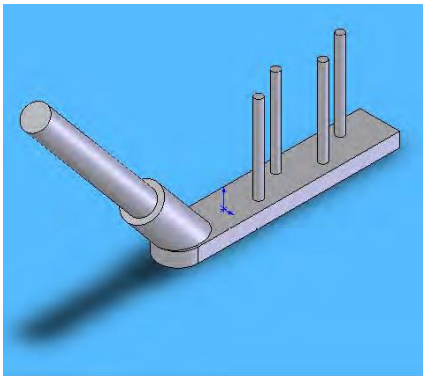


Figure 5. Designing of a dynamic screw bolt of a hip [6]

In [7] the description of a way and procedure of designing an internal fixation in the shape of a tile type "medially locking plate" (MLP) is described, which is used for treating femur fracture from its lateral side.

For designing process a 3D femur model is used. The positioning of a tile according to femur is defined by creating a datum plane. The datum planes were positioned in such a way that the sagittal plane was parallel to a planar approximation of the medial epicondylar surface and approximately tangent to the diaphyseal surface. The coronal plane was then position 90° to the sagittal plane rotated about a linear approximation of the diaphyseal axis (Figure 6).

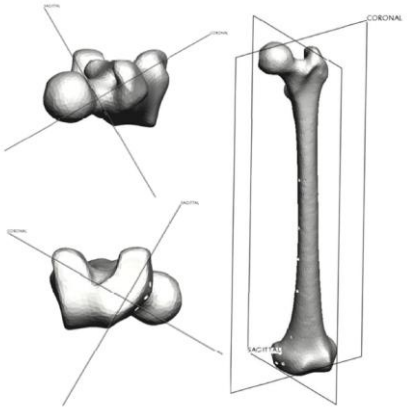


Figure 6. Creating a plane for drawing on a bone [7]

Next, a medial sketch of the MLP was created on the sagittal plane such that diaphyseal a d condylar shape matched that of the average femur (Figure 7).

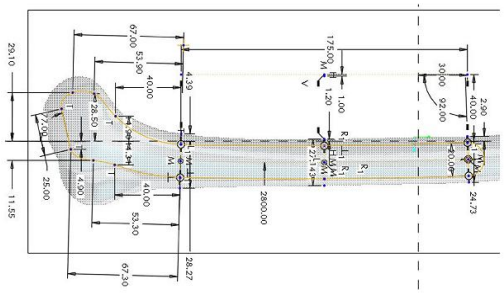


Figure 7. Creating the contour of an internal fixation [7]

The medial sketch was then extruded in both normal directions so that the extrusion intersected the femur at all points within the cross section and extended at least 1 cm beyond the most medial point on the medial epicondylar surface (Figure 8).

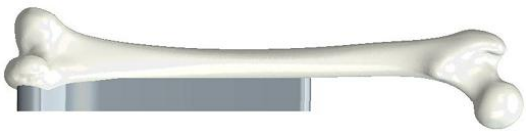


Figure 8. Extruding the contour of an internal fixation [7]

A sketch of the anterior profile – with an initial plate thickness of 5.0 mm – was then created on the coronal plane having a contour that matched the average femur. The sketch was then infinitely extruded in both directions, removing material where it intersected the initial medial/lateral extrusion (Figure 9).

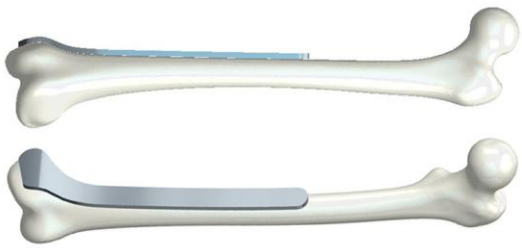


Figure 9. Extrusion cutting of the anterior profile [7]

Finally, the initial locations of the threaded interlocking screw holes were sketched and extruded - removing material when intersecting the concept model – from the sagittal plane (Figure 10). The final model of a tile optimized for structure integration is shown in Figure 11.

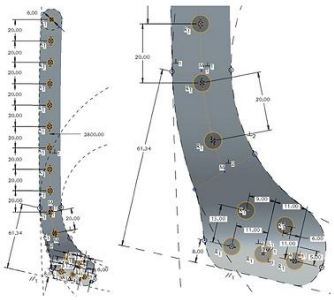


Figure 10. Creating the screw holes for the fixation screw bolts [7]

Figure 11. Final model of MLP internal fixation [7]

In [8] and [9] the method which is used for designing an internal fixation according to Mitkovic type TPL (tibia-plato-lateral) is shown. The suggested method is based on the application of MAF - Method of Anatomical Features and newly developed techniques for designing fixation supporting surfaces. The result of the application of this method is the parametrical fixation model whose shape and geometry could be changed with a change of parameters. With this approach it was allowed to change the shape of a fixation in order to adjust it to the patient's bone shape, in this case tibia, based on parameters values (dimensions) read from a suitable X-ray (for instance, a CT - Computer Tomography – scan) (Figure 12).

Figure 12. Fixation according to Mitkovic, type TPL created using the method described in [9]

IV. DESIGNING TECHNIQUE OF AN ANATOMICALLY ADJUSTED DYNAMIC INTERNAL FIXATION OF TIBIA ACCORDING TO MITKOVIC TYPE TPL

The principle of anatomical adaptability implies that the internal fixation with its contact surface fully corresponds to the surface of the part of a bone where the fracture is located. In this paper, we present the process of designing

an anatomically adjusted dynamic internal fixation of tibia according to Mitkovic type TPL, by using CATIA V5 software package.

Designing procedure is the following:

The parametrical 3D geometrical model (of tibia) made on the basis of patient's CT scan is used [10]. After that, a vertical plane not far from the lateral surface of tibia is created, which is placed opposite the contour of the fracture (Figure 13). Inside it, the contour of the proximal part of the fixation is drawn (Figure 14).

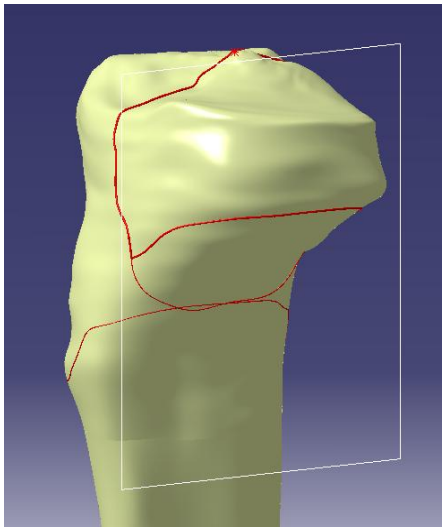


Figure 13. Creating the plane for contour drawing

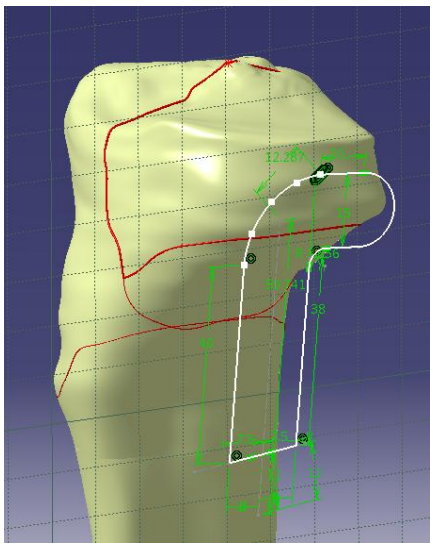


Figure 14. Creating the contour of the proximal part of the fixation

Following that, contour extrusion in the direction of the lateral tibia side is performed so as to ensure that extruded contour surface penetrates the bone surface (Figure 15).

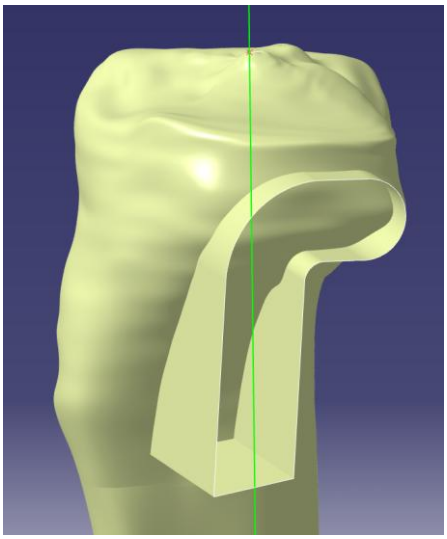


Figure 15. Extrusion of the contour and its penetration through the bone

The intersected closed contour of the fixation's internal side is created in this way (Figure 16). Inside of that contour, curve drawing of the 3D splines that follow the bone contour is performed (Figure 17).

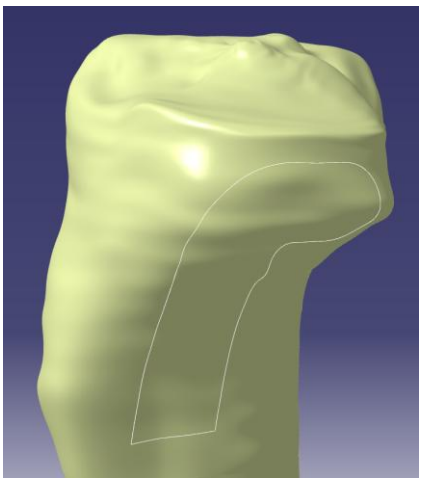


Figure 16. Creating the intersecting contour curve

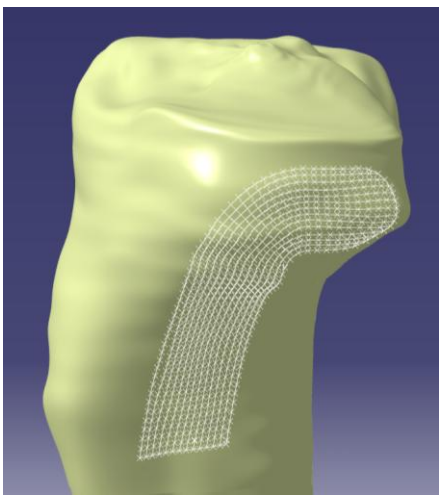


Figure 17. Creating 3D splines inside of the contour

After this step, the moment comes when all the surfaces are removed and the only surface left is the one

that actually presents the internal side of the fixation that is put directly on the bone (Figure 18a). Now this surface is extruded to transform it into a full model. With this process completed, we get a full 3D model of a proximal fixation part that is completely anatomically adjusted to the surface of the proximal part of tibia (Figure 18b).

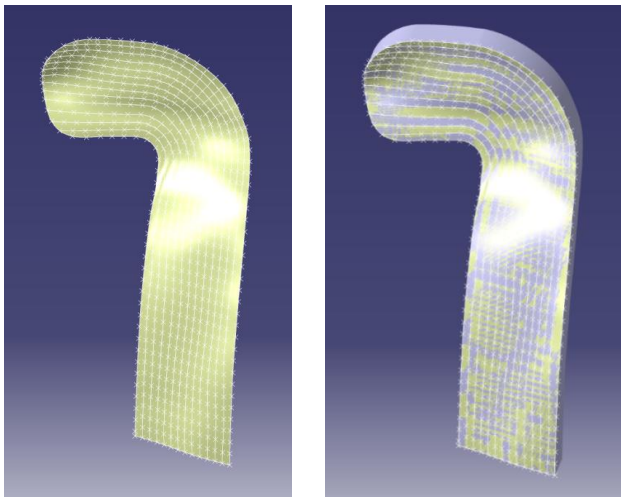


Figure 18. a) Creating a 3D surface; b) Creating a full model of proximal fixation part

The remaining parts of the internal dynamic fixation according to Mitkovic are made with the use of standard technical elements. What is characteristic for this part is a distal part of the fixation with two grooves used for the process of dynamization (Figure 19). In fact, the process of dynamization can be performed because of the lower groove with screw bolts, i.e. when the screw bolt from the upper groove, which enables previous deactivation of the whole process, is removed. In this way, a direct interaction between fractured bone fragments is created, in order to create new bone tissue and enable the bone to heal.

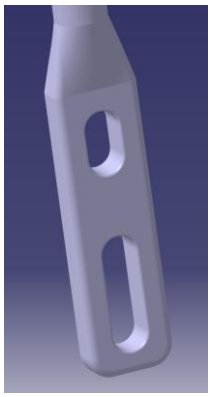


Figure 19. Distal fixation part with the grooves for dynamization

After the internal fixation has been shaped, the creation of the screw holes on the proximal fixation part is performed. According to the orthopedist request, an additional scheme of concentric circles with points for screw holes production is created (Figure 20). The process of screw holes creation is based on projection points and created tangent planes and is performed on the part of the proximal fixation surface (Figure 21).

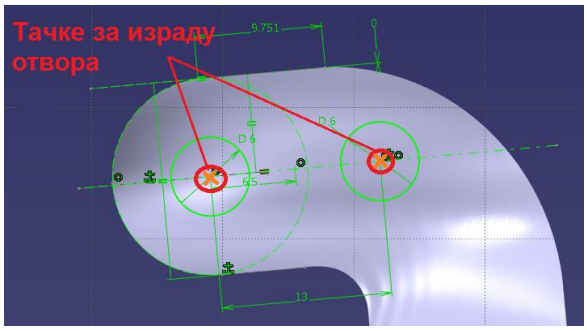


Figure 20. Creating an additional scheme with points for screw holes

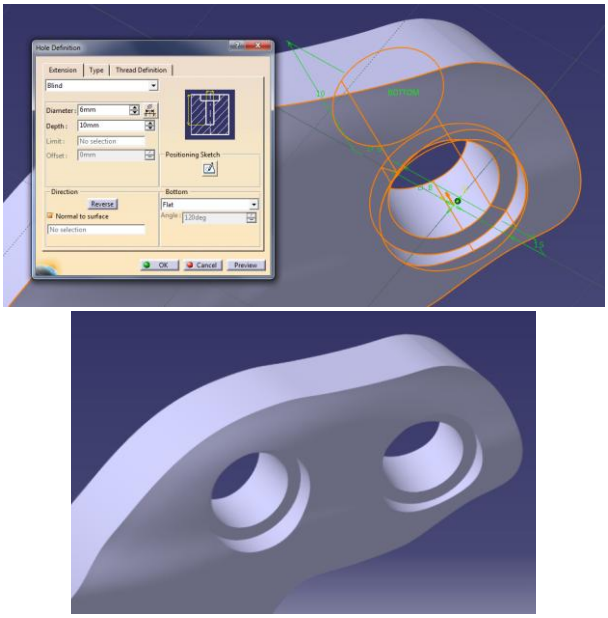


Figure 21. Creating screw holes

The final model of the internal dynamic fixation for tibia according to Mitkovic is shown in Figure 22.

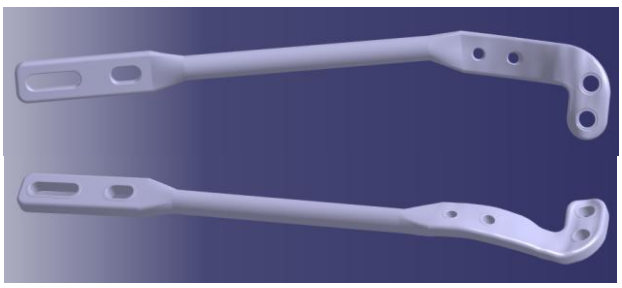


Figure 22. Final model of dynamic fixation according to Mitkovic

V. CONCLUSION

The method described in this paper presents the designing process of a 3D internal dynamic fixation model according to Mitkovic type TPL, whose internal surface lying on the bone is fully aligned with the bone surface. In this way, it ideally lies on the bone. With this model it is possible, based on a 3D model of bone and fracture, to

create a fixation for any fracture location. 3D fixation model is ideal for 3D printing or production using a CNC machine.

The method developed and described in this paper is applicable to many other implants of tile type and for any human bone. The requirement that must be fulfilled is a 3D bone model with imprinted fracture.

This method has significantly improved the technique for production of anatomically adjusted internal fixations.

ACKNOWLEDGMENT

This paper is a result of the project III41017, supported by the Ministry of Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] Mitkovic Milorad, Milenkovic S., Micic I., Mladenovic D., Mirkovic Milan, Results of the femur fractures treated with the new selfdynamisable internal fixator (SIF).. Eur J Trauma Emerg Surg. 2012 Apr;38(2):191-200

[2] Vitković, N., Milovanović, J., Korunović, N., Trajanović, M., Stojković, M., Mišić, D., Arsić, S. Software system for creation of human femur customized polygonal models, Computer Science and Information Systems, Vol. 10, No. 3, pp. 1473-1497, 2013

[3] D. Djenadić, M. Manić, D. Tanikić, S. Randjelović, P. Djekić, Analiza i prikaz vrsta fiksatora u medicini kao i metoda obrade materijala za izradu fiksatora, Vojnotehnički glasnik, Vol. 61, No. 2, pp. 123 - 139, 2013.

[4] http://www.synthes.com/MediaBin/International%20DATA/036.000.380.pdf. Accssed on 9 Jan 2015.

[5] Matthys R, Perren SM. Internal fixator for use in the mouse. Injury, Int. J. Care Injured (2009) 40S4, S103– S109

[6] Nooshin Sadeghi Taheri, Modelling and analysis of a dynamic hip screw: biomechanical analysis of a dynamic hip screw under different load conditions, Master thesis, Swinburne University of Technology Faculty of Engineering and Industrial Sciences, (2011), pp 57-58.

[7] Arnone Joshua. A comprehensive simulation-based methodology for the design and optimization of orthopaedic internal fixation implants, Ph. D., The Faculty of the Graduate School, University of Missouri-Columbia, 2011.

[8] Vitković, N., Veselinović, M., Mišić, D., Manić, M., Trajanović, M., Mitković, M., Geometrical models of human bones and implants, and their usage in application for preoperative planning in orthopedics, 11th International Scientific Conference MMA 2012 - Advanced Production Technologies, Novi Sad, 2012, pp 539-542

[9] Dalibor M. Stevanović, Nikola M. Vitković, Marko M. Veselinović, Miroslav D. Trajanović, Miodrag T. Manić, Milorad B. Mitković, Parametrization of internal fixator by Mitkovic, International Working Conference ''Total Quality Management – Advanced and Intelligent Approaches'', 4th – 7th June, 2013., Belgrade, Serbia, pp 541-544

[10] Vidosav Majstorovic, Miroslav Trajanovic, Nikola Vitkovic, Milos Stojkovic, Reverse engineering of human bones by using method of anatomical features, CIRP Annals - Manufacturing Technology 62 (2013) 167–170 (M21, IF 2,251)

Methods for assessment of cognitive workload in driving tasks

Kristina Stojmenova and Jaka Sodnik
University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia
kristina.stojmenova@fe.uni-lj.si
jaka.sodnik@fe.uni-lj.si

***Abstract*— In this paper, we explain the concept of mental or cognitive workload and review the most common methods and procedures for its assessment. We focus primarily on driving tasks and interaction with various in-vehicle devices and systems. Safety is one of the most important human needs and consequentially also the primary concern of the automotive industry when introducing new in-vehicle information systems (IVIS), global positioning systems (GPS), interactive displays, etc. Since the human brain and its resources are limited, the primary task of driving can be seriously challenged when secondary tasks are performed simultaneously. Several different methods have been proposed for direct and indirect measurement of the driver's cognitive workload and for the detection of its potential overload. We report briefly also on two user studies performed in our driving simulator, which illustrate the importance of correctly assessing cognitive workload in the process of evaluating new in-vehicle user interfaces.**

I. INTRODUCTION

In all developed countries the majority of adult population owns a driving licence and participates in traffic on daily basis. With the increased number of drivers and vehicles, attention to driving and everything that influences it, has increased significantly. Although the safety of cars increases every year, they are also equipped with variety of electronic devices aiming to assist the driver in the driving process, enabling navigation, communication and entertainment services. These devices can distract the driver, negatively affect his or her responsiveness and cause significant amount of unwanted workload.

Workload is generally defined as the amount of work an individual has to do [1]. The term "workload" covers a broad spectrum of human activities and corresponding tasks while the term "mental workload" focuses only on demands imposed on the human's limited mental resources [2][3]. In both cases there is always a difference between the individual's subjective perception of workload and the actual amount of work. The so called overload of workload arises when the amount of human activity and requests for various physical and mental resources surpasses the amount of the available processing resources (of human brain or more precisely of the correspondent part of the brain).

Attention on the other hand, is defined as concentration of awareness to a specific source of information, whereas distraction is the diversion of attention from one source of information to one or several other sources [4]. Humans have multiple but limited amounts of attention and processing resources available at any given time [5]. Different tasks can use different attention resources or share them. If several simultaneously performed tasks rely on the same source they usually interfere with each other and seriously compete for that resource (e.g. two separate visual tasks) [6]. In that case the brain cannot process all the information presented and the performance is significantly affected. In a lot of real life situations this can be hazardous especially if the overlooked information regards safety, health or security issues.

In the past, a mental workload and measurement of mental over-workload were investigated and considered primarily in relation to design and operation of aircrafts and onboard systems in aircrafts. The mental workload of pilots can sometimes be critically increased due to variety of functionalities and features of these onboard systems and a variety of tasks each pilot has to preform simultaneously. Similarly, the available functionalities of In-Vehicle Information Systems (IVIS) increase rapidly, expecting a driver to perform high number of tasks simultaneously. However, a human brain can process only a limited amount of information. When the amount of mental workload surpasses the amount of mental capability, the surpassed amount of information is missed. Unfortunately the brain, more or less, randomly selects the information that is going to be ignored without letting a human consciousness to choose what is more and what is less important. While in many situations this may not be so important, for drivers of vehicles it can be dangerous and critical.

Mental workload can be measured directly by measuring the cerebral activity of human brain and indirectly with a number of different experimental methods. In this paper we review several methods for measuring mental workload of drivers while operating a vehicle and performing different secondary tasks (e.g. using on-board computer or infotainment system in a vehicle). The reviewed methods are divided into four major categories:

- direct psychophysiological measurements,
- measurement of ocular activity,
- methods based on measurement of response time and
- subjective measurements based on questionnaires for self-evaluation.

We also briefly report on two examples of cognitive workload measurements in a driving simulator revealing its importance in evaluation of novel in-vehicle systems and display technologies.

II. PSYCHOPHYSIOLOGICAL MEASUREMENTS

Due to the multifaceted nature of the complex mental demands in in-vehicle interaction multiple measures are required. One way of measuring mental workload directly is collecting the time-varying spatial potential distribution over the scalp produced by the cortical brain activity. These measurements can be performed with the electroencephalogram (EEG) or its magnetic counterpart magnetoencephalography (MEG) [7]. The brain activity is recognized whenever a potential difference appears between the electrode with an active neural signal and the electrode that is placed in an inactive surface to serve as a reference point. Based on the type of the activity that provokes the signal (spontaneous or event-related), the obtained results are divided in two categories. First group of signals correspond to spontaneous activities which can usually be detected in the frequency range between 1 Hz and 100 Hz. The second group of signals corresponds to various sensory, cognitive or motor events, which can be detected as event-related potential (ERP) in the brain. The ERP measurements always contain some noise which comes from other bio-signals in the brain and various electromagnetic interferences in the environment. Based on the assumption that the noise can be approximated by a zero-mean Gaussian random process, the resulting signal-to-noise ratio (SNR) can be significantly improved by averaging several ERP measurements.

Another measurement of the activity of the central nervous system is based on the analysis of variation of the potential distribution across the eye. The latter reveals the information about the eye blinks and the eyes movements (electrooculogram, EOG). There are also several other methods for measuring cognitive workload based on human's ocular activity. They are discussed in a separate chapter of this paper.

Increase or decrease of mental workload can be obtained also by monitoring the variation of the heart rate with the electrocardiogram (EKG), or by monitoring the variation of the galvanic skin responses (GSR) as well. Variation of the heart rate is recorded by skin electrodes that measure electrical activity caused by depolarization and polarization of the heart muscle, where increased mental workload leads to an increased cardiovascular activity and vice versa.

III. MEASURMENTS OF OCULAR ACTIVITY

Many ocular activity variables have been tested and validated to be correlated with mental workload, such as pupil diameter, blink rate, blink duration, blink frequency, fixation duration, fixation frequency and saccadic extent. All of these have been used to measure mental workload in vehicles. Experiments have shown that the physiological and performance measures and the remote eye tracking might provide reliable driver cognitive load estimation, especially in simulators [8].

A. Pupil Dilation

Pupil diameter is a physiological measure of cognitive workload. The magnitude of pupil dilation can be described as a function of processing mental workload required to perform a given task [9]. When facing an increased visual or mental workload people's pupils tend to increase in diameter. This phenomenon is called the Task Evoked Pupillary Response. Even though it can be measured with most of video based eye tracking systems, setting the system and providing accurate measurements is not effortless. This is mainly due to the visual interference that can appear and is not directly connected to the task. The pupil is very light sensitive and can react to every light change in the environment, such as for example, a vehicle driving in a dynamic road with random lighting. The size of the pupil, as seen by the eye tracking system, depends on the person's gaze angle. This issue is an important source of error in the measurement of pupil dilation. However, these technical issues can always be omitted and fixed. The procedure of measuring mental workload by observing changes in pupil diameter can therefore, be considered as a reliable indicator of changes in the mental workload.

B. Blink Duration and Blink Frequency

Eye blink parameters such as blink duration and blink frequency have been associated with level of drowsiness and information processing. Blink frequency has been reported to increase with greater workload in a way that the number of blinks increases as a function of time in ongoing tasks [10]. Blink duration, however, decreases when the driver is at the beginning of a new task and increases when the driver experiences fatigue or drowsiness. These conditions decrease driver's mental capabilities and increase overall cognitive workload.

Although all studies confirm that there is a clear correlation between these parameters and the mental workload, the actual results suggesting which parameter is more suitable differ drastically. Fakuda et al. found a correlation between pupil diameters and blink frequency which does not depend on task completion time or amount of information which needs to be processed [10]. The latter suggest that blink frequency can be considered as accurate indicator as the pupil dilation but much easier to detect and measure. Benedetto et al. on the other hand, suggest that blink duration is even more sensitive and reliable indicator of driver's visual workload as the before mentioned blink frequency [11].

IV. DETECTION RESPONSE TASK MEASUREMENTS

Detection response task (DRT) or Peripheral detection task (PDT) is a method for measuring the amount of driver's mental workload with the use of a secondary task distraction. DRT has been used in simulator and driving studies in recent years to assess changes in workload during driving, and to assess workload and distraction caused by in-vehicle information systems [12]. The most widespread version of this method consists of a red dotted light that is placed in front of the driver and a physical button placed on the steering wheel as seen in Fig. 1. The red light emitter turns on randomly every 2 to 5 seconds and emits for 1-2 seconds unless it is turned off earlier by the driver with the remote button. The task of the driver is to press the button each time he or she spots the red light. If a new interval starts without the driver pressing the button for the previous one, it counts as miss. The system keeps track of the time needed for the driver to respond to

the visual stimuli and all the missed targets. It also counts as a miss, if the driver presses the button more than once for one light interval. All this data is then collected and time-aligned with primary and secondary tasks. The longer response time and the higher number of missed targets signify greater mental workload of the driver.

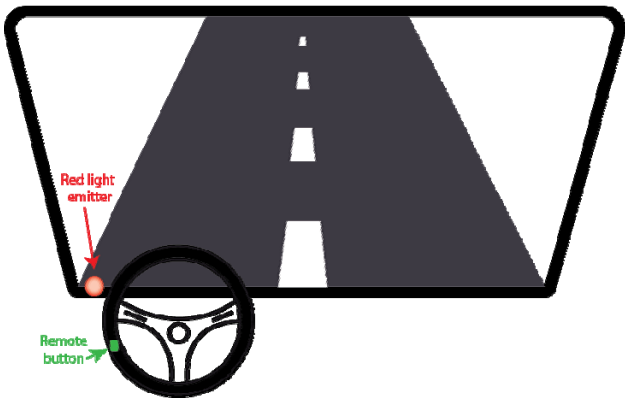


Figure 1. DRT method

Krause M. and Conti A. proposed the DRT measurement device as a simple mobile application which they named MDT [13]. It is available for Android devices and it is free of charge. The MDT works in the same way as DRT, the only difference is that the remote button and the light emitter are now both presented by the telephone's screen (see Fig. 2), which turns red every 2-5 seconds and can be turned off by pressing directly on the screen.

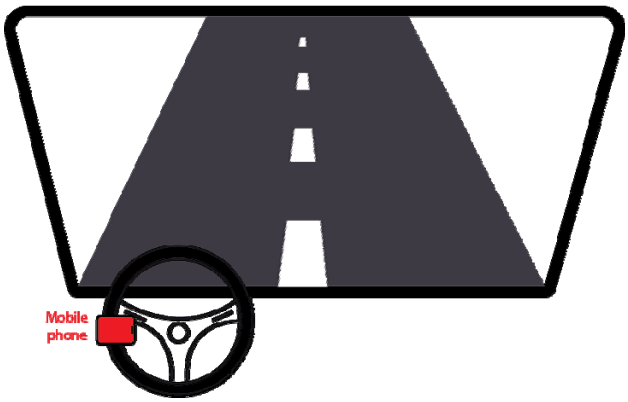


Figure 2. MDT method

Bruyas and Dumont split DRT methods into three categories, based on the stimulus used and its physical placement relatively to the driver [14]. The first method is called Head-mounted DRT (HDRT) since the small diode emitting red light is mounted directly on a driver's head, right in front of the eye (e.g. at the distance of approx. 10 cm). In the other method called Remote DRT (RDRT) the light source is placed in the lower right corner of the driver's windshield. While these two versions provide visual distractions, the third method called Tactile DRT (TDRT) uses tactile stimuli. A small vibrating sensor sends vibration impulses directly to the driver's skin.

The Detection Response Task measurement is currently being discussed in the ISO working group TC22/SC13/WG8 as the basis of a standard to assess the effect of cognitive workload on driver's attention [14].

V. SUBJECTIVE MEASUREMENTS

By subjective measurements we refer to questionnaire based self-evaluation procedures where test subjects are asked to answer different questions regarding their subjective perception of cognitive workload. These questionnaires are always completed after participation in various tasks and experiment conditions.

A. *NASA-TLX*

The Nasa Task Load (NASA-TLX) is a multi-dimensional rating procedure that provides an overall workload score based on a weighted average of ratings on six subscales [15]:

- Mental Demands,
- Physical Demands,
- Temporal Demands,
- Own Performance,
- Effort and
- Frustration.

Initially there were nine categories including also fatigue, stress and frustration. These three categories were later abandoned as they were found irrelevant for the final assessment of perceived cognitive workload.

The NASA-TLX procedure consists of two parts or two sets of questions (in a paper or an electronic version):

- rating of workload in different categories and
- estimation of weights for individual subscales.

The first part of the procedure requires the test subject to rate each of the subscales (e.g. magnitude of load) on how much is it present and important for a given task. Individual subscales are rated on a scale from "low" to "high" divided into 20 equal intervals (e.g. the final score ranges from 0 to 20).

Estimations of weights is performed by comparing different categories (subscales) pair-wise, forming 15 different comparisons. Whenever an individual subscale (e.g. the source of workload) is selected to be a greater contributor to the anticipated workload for a specific task its counter is incremented by one. The final number of counts represents the weight for individual subscale and ranges from 0 to 5. The overall workload for a selected task is than calculated by summing products of ratings and corresponding weights. Additionally the sum is divided by 15 (the sum of the 15 paired weights) to normalize the final score.

NASA-TLX has been used in various experiments for estimating cognitive workload of drivers offering a reliable comparison of different experiment conditions and tasks ([6], [8], [11], [12], [19], [20], etc.).

B. *DALI*

The Driving Activity Load Index (DALI) was developed with as a subset of NASA-TLX. While the latter was initially developed to measure pilot's cognitive load, DALI on the other hand is a reversed version which

has been adapted for accessing driver's cognitive load [16]. It also consists of 6 subscales:

- Effort of attention,
- Visual demand,
- Auditory demand,
- Temporal demand,
- Interface and
- Situational stress.

These changes were applied since some of the subscales in TLX cannot be entirely correlated with the task of driving (e.g. "physical demand" or "performance"). Pauzé discusses that "physical demand" for example cannot be considered relevant for driving a car since today's cars demand only negligible physical effort in order to be operated efficiently[17]. She also talks about how the "performance" estimation may vary from person to person due to its self-esteem, motivation to fit the expected standards and other similar factors that are not directly related to the mental workload.

The original NASA-TLX is still the simplest and the most often used method for the estimation of cognitive workload also for driving tasks and simulator studies. Consequentially, it is also the most commonly cited method enabling different researchers to directly compare their results. In the following chapter we report briefly on two of our experiments in driving simulator where NASA-TLX was used for estimation of cognitive workload and showed some significant differences between different display technologies in vehicles.

VI. Examples of using NASA TLX for the evaluation of auditory and visual interfaces in vehicles

The main research goal of the first study was to establish if a pure auditory interface in a vehicle can perform better than a classical visual interface displayed on a Head-down Display (HDD) [18]. We intended to find out if it is less distracting for a driver by measuring cognitive workload of drivers and evaluating their driving performance (e.g. counting driving errors and potential dangerous situations on the road, etc.). The user study took place in a driving simulator, consisting of a large projection screen, steering wheel, pedals and the in-build IVIS supporting a variety of tasks related to communication, navigation and multimedia. The IVIS was operated through a custom-made interaction device attached to the steering wheel, which enabled the drivers to enter commands and select functions while holding the steering wheel. The feedback or the output of the system was based on three different display technologies:

- visual HDD mounted on the dashboard,
- mono auditory interface played through a speaker and
- spatial auditory interface played through a 7.1 speaker surround system.

The three technologies were used separately in three isolated experiment conditions. The simulator setup is shown on Fig. 3.

Figure 3. Driving simulator used for evaluation of spatial auditory interfaces and HDD [18]

An important research question was also if the use of spatial audio and multiple simultaneous sounds can increase the efficiency of such auditory interfaces (e.g. by comparing mono and spatial audio). Users were asked to perform different secondary tasks of varying complexity while driving a simulated vehicle (e.g. find a song in the multimedia device and play it, call somebody from the list of contacts, send a short txt message to somebody, etc.) Their performance was evaluated through the measurement of task completion times, counting of driving errors and anomalies and also NASA TLX test. The latter was included to compare individual interfaces by the amount of cognitive workload they put on a driver.

Both auditory interfaces proved to be significantly safer as the HDD resulting in much lower number of driving errors and hazardous situations on the road. They also proved to be equally fast and effective for performing secondary tasks while driving and causing less cognitive workload to the driver. No significant differences were found between the two auditory interfaces in terms of task completion time or driving errors. On the other hand, the results of NASA TLX test reviled some important differences between the two auditory interfaces (see Fig. 4).

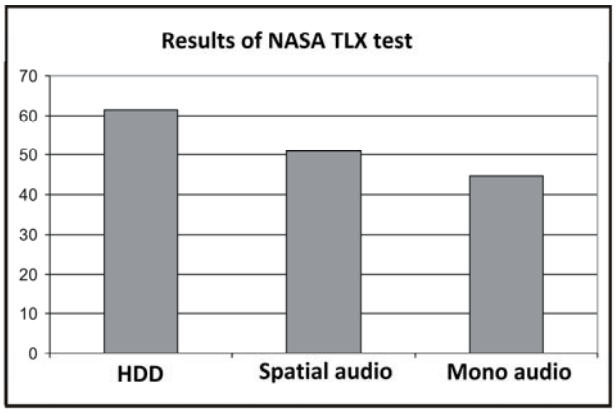


Figure 4. Overall scores of NASA TLX test comparing three experiment conditions [18]

The use of spatial sound and multiple simultaneous sounds resulted in significantly higher cognitive workload. Majority of test subjects substantiated this by clearly explaining that perception of spatial sound in such form is inappropriate and too much mentally demanding for in-vehicle environments.

In the second study we changed the HDD to a Head-up Display (HUD) where output of IVIS was projected directly to the windshield [19]. In this way, it is much easier to read the content and less "eyes off the road" situations occur. The HUD was compared to the mono auditory interface and to a multimodal display (e.g. using HUD and audio output simultaneously). The main research question was how the change of HDD to HUD will reflect in task completion time, safety of driving and cognitive workload. We were primarily interested if a visual display projected as a HUD can be comparable to pure auditory display. Beside the display technologies, the simulator software was also changed in this study resulting in an even more realistic driving experience (see Fig. 5).

Figure 5. Driving simulator used for evaluation of auditory, visual (HUD) and multimodal interfaces [19]

In comparison to the previous study, the results in this case did not show any significant differences between the HUD and the auditory interface. The visual interface did not show to be any less safe than the auditory interface, resulting in a comparable number of driving errors and anomalies. The NASA TLX scores on the other hand, indicated some significant differences between the interfaces. In this study, the estimated cognitive workload caused by the auditory interface proved to be higher than the workload caused by the HUD or by the multimodal interface (see Fig. 6). The latter proves that HUD and multimodal displays should be the next step in the field of in-vehicle display technology providing better safety and lower cognitive workload.

These are just two examples illustrating the importance of correct estimation of cognitive workload in driving tasks. The NASA TLX tests showed some important differences between the evaluated displays and technologies which were not registered through the measurement of other objective parameters (e.g. task completion times and driving errors).

Recently we have started using the DRT method in our research, also for the assessment of cognitive workload in various driving tasks. We are currently conducting an extensive research study in which we are comparing the newly proposed audio version of the DRT to the well reported and commonly used visual and tactile versions of the DRT test.

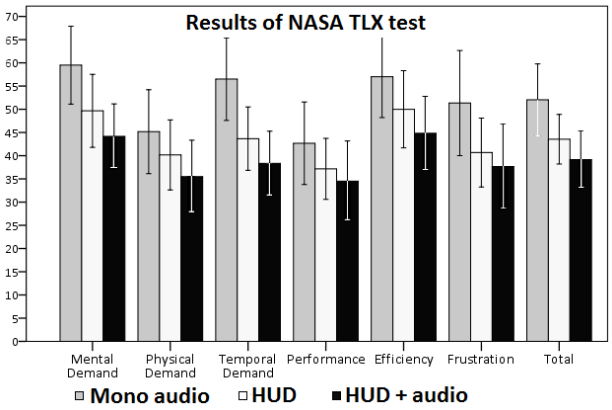


Figure 6. NASA TLX scores of individual subscales comparing three experiment conditions [19]

VII. DISSCUSSION AND CONCLUSION

Nowadays mobility represents an essential part of our lives and majority of the tasks that we used to perform in the office are being done also while walking, running, cycling and even driving a car. All these primary and secondary activities compete for a limited capacity of our cognitive and mental resources. If the cognitive workload of numerous simultaneous activities exceeds the available mental resources, our performance drops significantly and leads to dangerous and unpredicted situations. It is therefore very important to correctly estimate and measure the cognitive workload caused by individual activities and by the use of specific electronic devices and user interfaces. In this paper, we summarized the most relevant methods for assessment of cognitive workload of drivers when driving a car and performing various secondary tasks in a vehicle.

The psychophysiological measures represent a group of the most advanced methods for assessment of cognitive workload, since they require a set of very expensive and complex equipment as well as well-trained professionals capable of recording and interpreting the brain signals. They offer continuous observation with high time resolution and collection of data without disturbance and intrusion into primary tasks [20].

Similarly, the measurements of ocular activity could also represent an accurate and very objective measurement of cognitive workload in various environments and tasks. However, due to the specifics of the driving environment and the complexity as well as high price of the required equipment they are not as commonly used as the DRT methods.

DRT methods are rapidly gaining on their popularity and are being more and more used also for observing in-vehicle activities and tasks. However, the visual version of DRT test itself causes a significant amount of workload and high demand specifically for visual resources. It is therefore unsuitable for complex visually-demanding secondary tasks and should be replaced with the tactile or perhaps auditory version of the same test (e.g. the use of tactile or auditory stimulus).

The subjective questionnaire-based procedures are still most widely used and cited methods for assessing cognitive workload in driving tasks. The results obtained with NASA TLX test can, for example, be directly compared to the biggest number of results of other similar studies and researches. Another major benefit of NASA TLX is also its simplicity, since it requires no specific or sophisticated electronic equipment. On the other hand, the problem with subjective measures is that it is always performed post-hoc and does not measure time-varying qualities. The answers in the questionnaires are for example strongly influenced by events towards the end of the task (e.g. they are more related to the latest events instead of all events which occurred throughout the task) [21].

In general, it is hard to estimate which method is more suitable than the other. This is because the outputs of various methods are not comparable and do not have the same reference parameters. Some measure only the visual workload while the others measure the total workload without knowing which resource centre is actually used for successful performance of various tasks.

Finally, it is important to point out that the mental workload in driving tasks should not be to low either. Small amount of mental demand can cause the driver to turn from an active to a passive participant in the process, making the driving process monotonous and dull (e.g. when driving in semi-autonomous or autonomous vehicle). This can quickly lead to sleepiness and lack of motivation. The final result and mental state can be as dangerous as a mental overload.

In the future, it will be therefore, very important to constantly estimate the level of cognitive workload of drivers and sustain it in the predefined and safe range.

REFERENCES

[1] Workload, From Wikipedia. Available from: http://en.wikipedia.org/wiki/Workload.

[2] Wickens, C.D. and Hollands, J.G., "Engineering Psychology and Human Performance", Prentice Hall, Saddle River, New York, 2000.

[3] N. Meshkati and P.A. Hancock (editors), "Human mental workload", *Elsever*, p.8, September 2008.

[4] James, W., "The Principle of Psychology", New York, NY, USA: Holt, 1890.

[5] Navon, D. and Gopher, D., "On the economy of the human processing system," *Psychological Review*, vol. 86, no. 3, 1979, pp. 214–255.

[6] Harms, L. and Patten, C., "Measuring distraction from navigation instructions in professional drivers—a field study with the peripheral detection task". In: *Proceedings from 9th International Conference of Vision in Vehicles*, Brisbane. Elsevier, Exeter, 2001.

[7] Borghinia G., Astolfia L., Vecchiato G., Mattiaa D., and Babiloni F., "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness", *Neuroscience and Biobehavioral Reviews* vol. 44 pp.58–75, 2014

[8] Palinko O., Kun A. L., Shyrokov A., and Heeman P., "Estimating cognitive load using remote eye tracking in a driving simulator". In Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, 2010, pp.141-144.

[9] Beatty J., "Task-evoked pupillary responses, processing load, and the structure of processing resources". In Psychological Bulletin, vol.91 (2), 1982, pp.276–292.

[10] Fakuda K., Stern J. A., Brown T. B. and Russo M. B., "Cognition, Blinks, Eye-Movements, and Pupillary Movements During Performance of a Running Memory Task". In: Aviat Space Environ Med, vol. 76, 2005, pp.75-85.

[11] Benedetto S., Pedrotti M., Minin L., Baccino T., Re A. and Montanari R., "Driver workload and eye blink duration". In: Transportation Research Part F, vol. 14, 2011, pp. 199-208.

[12] Jahn G., Oehme A., Krems J. F. and Gelau C., "Peripheral detection as a workload measure in driving: effects of traffic complexity and route guidance system use in a driving study", Transportation Research Part F:Traffic Psychology and Behaviour, vol. 8, May 2006, pp.255-275.

[13] Krause M. and Conti A., "Mobile detection taks". Available from: http://www.lfe.mw.tum.de/en/research/open-source/mdt/.

[14] Bruyas M. P. and Dumont L., "Sensitivity of detection response task (DRT) to the driving demand and task dificulty". In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assesment, Training, and Vehicle Design*, 2013.

[15] NASA TASK LOAD INDEX (TLX) Paper and Pencil Manual. Available from: http://humansystems.arc.nasa.gov/.

[16] Pauzié A. And Pachiaudi G., "Subjective evaluation of the mental workload in the driving context". In: Traffic and Transport Psychology: Theory and Application, 1997, pp. 173-182.

[17] Pauzié A, "Evaluating driver mental workload using the driving activity load index (DALI)". In Procedings European Conference on Human Centred Design for Intelligent Transport Systems, 2008, pp. 67-77.

[18] Sodnik, J., Dicke, C., Tomažič, S., Billinghurst, M., "A user study of auditory versus visual interfaces for use while driving". In: International Journal of Human-Computer Studies vol. 66(5), 2008, pp. 318-332.

[19] Jakus, G., Dicke, C., Sodnik, J., "A user study of auditory, head-up and multi-modal displays in vehicles". In: Applied ergonomics, vol. 46, 2015, pp. 184-192.

[20] Wilson, G.F. and Russell, C.A., "Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding". In: Human Factors, vol. 49, 2007, pp. 1005-1018.

[21] Insko B.E., "Measuring presence: Subjective, behavioral and physiological methods". In: Being There: Concepts, effects and measurement of user presence in synthetic environments, 2003.

On the Runtime Models for Complex, Distributed and Aware Systems

Milan Zdravković*, Miroslav Trajanović*
* Laboratory for Intelligent Production Systems (LIPS),
Faculty of Mechanical Engineering, University of Niš, Niš, Serbia

milan.zdravkovic@gmail.com, miroslav.trajanovic@masfak.ni.ac.rs

***Abstract* – Recent developments in the area of Internet of Things increase the pressure on the feasibility of current architectures of the Enterprise Information Systems (EIS), in terms of their complexity, flexibility and interoperability in a pervasive computing world. The fact that EISs are today hosted by the growing diversity of platforms and devices, urges as to consider new concepts that would take into account rapid deployment and setup in any circumstances. This paper presents the discussion of model-driven architectures and proposes the concept of EIS design that is ontology-driven, persistence-neutral, runtime-model-based. These concepts are to some extent demonstrated in the case of OntoApp tool for ontology scaffolding.**

I. INTRODUCTION

The emergence of the ubiquitous computing technologies, such as Wireless Sensor Networks (WSN), Cyber-physical Systems (CPS) [1], Internet-of-Things (IoT) [2] and Future Internet Enterprise Systems (FinES) [3] is posing the new challenges to the traditional body-of-knowledge and practice in the design and development of Enterprise Information Systems (EIS). The increasing diversity and multiplicity of platforms where EISs are operating (taking into account different devices, e.g. sensors, processing devices and actuators) and lack of common, unifying standards and theories bring the distributed, federated, adaptable architectures, with so-called self-* properties into the focus of EIS developers and users.

The huge number of identifiable devices, used also on a sharing basis, is expected to become a commodity in the future, also providing a technology tool for emergence of so-called "sensing enterprise" [4]. Such diversity will pose tremendous challenges related to the interoperability issues. The new technological landscape, provided by the Future Internet systems will thus establish interoperability problems as critical and possibly consider the interoperability as an inherent capability of the future information systems.

In the recently submitted position paper, IFAC TC5.3 Technical Committee for Enterprise Integration and Networking of the International Federation for Automatic Control addressed the several research challenges of the systems interoperability, in attempt to define the future research directions towards so-called Next Generation Enterprise Information System (NG EIS). The following properties have been identified as critical for NG EIS: omnipresence, model-driven architecture, openness, dynamic configurability, multiplicity of identities, awareness/inclusive sensing and computational flexibility. Based on these properties, a generic, abstract architecture has been proposed, as illustrated on Fig.1.

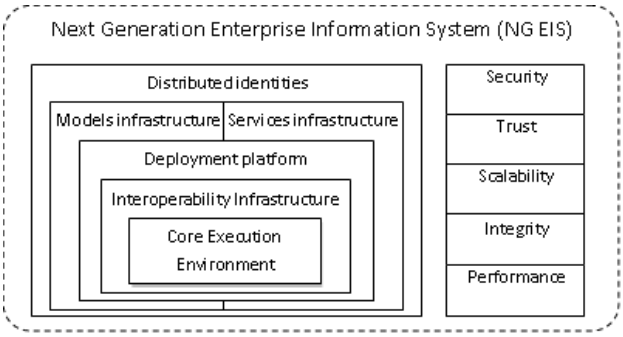


Fig.1 Abstract architecture of NG EIS, according to IFAC TC5.3

The main objective of the above proposal is to promote the research related to removing the complexity from the EIS design. In the ideal scenario, NG EIS will become a software shell, a core execution environment with the integrated interoperability infrastructure. Such an environment is foreseen as a highly flexible and scalable, deployable on any and every platform, using the external models and services infrastructure, exclusively or on a sharing basis. Finally, NG EIS will be highly extrovert system, represented by its multiple identities, e.g. as agents, UIs, services, etc.

In this position paper, we further build on that vision proposal by presenting the initial discussion on the possibilities to use conventional Model-Driven Engineering (MDE) tools and practices as enablers of NG EISs. When considering the objective of the technical unification of the core execution environments of NG EISs, we foresee that these environments will be driven by so-called runtime models. These runtime models will be the formal descriptions of the different aspects of the NG EIS design and operation, including data and information layer, business logics and UI. Finally, the objective of the work behind this paper is not to propose new formal models which will act as runtime models, but rather to rely on the vast number of existing domain and application ontologies.

II. THEORETICAL BACKGROUND

The complete software development cycle can be viewed as a form of implicit modeling. For example, the problem analysis and identification of the solution is based on the experience of software architects, where this

experience can be in fact considered as a model of knowledge. Then, this model of knowledge is specialized to the mental schemas of the architects.

Similar can be said also for the coding process. A code is also an implicit model, because it is built based on the specific language abstractions (syntax), software patterns and mental schemas of the coders who combine the syntax with the software patterns and experience to produce the code [5]. In this process, a gap between the problem and implementation definitions appears when "a developer implements software solutions to the problems by using abstractions that are at a lower level than those used to express the problem." [5]

The development process issues that contribute to the gap mentioned above are commonly addressed by Model-Driven Engineering (MDE) practices. MDE is a software development approach in which abstract models of software systems are systematically transformed to their specific implementations. MDE is driven and motivated by the growing complexity of software, which additionally increases the gap between the problem-level software abstractions (e.g. requirements) and its implementation.

Today, MDE approaches and practices are commonly addressed by using a Model Driven Architecture (MDA) [6]. MDA is a framework of MDE standards, launched and maintained by Object Management Group (OMG). It distinguishes between computation independent (CIM), platform independent (PIM) and platform specific (PSM) models. Main pillars of MDA are Meta Object Facility (MOF) language for defining the abstract syntax of modeling languages [7], UML [8] and Query, View, Transformation standard (QVT) for specifying PIM to PSM transformations [9].

Existing MDE tools and practices assume fixed viewpoints to one system. Now, each of these viewpoints needs also to consider the multiple identities of one system. The separation of concerns (e.g. functional, security, privacy, performance, etc.) in MDE approach needs to consider these identities.

Model transformations are considered as some of the key enablers for system interoperability. Despite the extensive results in the area of meta-modeling (e.g. MOF), the foundation for specifying transformations between models has not been built yet [10].

A. Formal Specification Techniques (FST)

One of the main problems of the current models is a lack of validation tools. Typically, the models of complex information systems are extensively large and in general, there exist no tools for querying and navigating them. More important, there exist no tools for their analysis. Thus, it becomes very difficult to maintain their consistency.

Formal Specification Techniques (FST) aim at restricting the modeling viewpoints, with objective to provide analysis, transformation and generation tools. A common approach is to translate a modeling view (e.g. a UML class model) to a form that can be analyzed using a particular formal technique [11]. This analysis can involve the consistency checking (for example, the relationships between the occurrences of the same software artifact in different viewpoints), completeness and dependability. Thus, reasoning on the formal specification of one system can be further used to prove that all actions will result in a discrete set of states, that some system properties are bounded, that error states are unreachable, etc.

Use of FST dates from the late seventies. Abrial et al [12] have proposed Z notation – a formal specification language for describing and modeling computing systems, based on Zermelo-Fraenkel set theory. Alloy has been developed [13] as a language for describing structural properties and their automatic semantic analysis. It is associated with an analyzer tool [14].

FST aims at facilitating so-called transformational programming or program transformation. The latter refers to an operation which transforms one computer program to another, which is "semantically equivalent to the original, relative to a particular formal semantics" [15]. The transformations are carried out incrementally, in manageable, controlled transformation steps which guarantee that the final software product will meet the initial specification. Although first concepts of program transformation has been defined in early seventies [16], the first exhaustive methodology was defined in scope of the Munich project CIP (computer-aided intuition-guided programming). That research included the "design of a wide-spectrum language specifically tailored to the needs of transformational programming, the construction of a transformation system to support the methodology, and the study of transformation rules and other methodological issues." [17].

Although current FST techniques are considered as self-sufficient, many authors addressed the problem of transforming widely accepted UML models to formal specification languages. Precise semantic characterizations of Object-Oriented modeling concepts have been introduced in 1997 [18]. The different tools have been developed to facilitate transformations of UML annotated class diagrams to complete Z [19] or Alloy [20] specifications. Csertdn et al [21] developed a transformation-based verification and validation environment for improving the quality of systems designed using the UML by automatically checking consistency, completeness, and dependability requirements.

1) Formal specification of business logic

Besides formal verification and validation, FST could also have significant role in semantically annotating the IS architecture, as well as the code. This can be achieved by correlating the specific FST formalisms to the lower level semantics of the domain ontologies, as well as the higher level semantics which is using generic, IS concepts, but is not bound to the specific domain.

For example, high level semantics is sometimes used to describe the businesses, e.g. by taking process perspective (BPEL). One of the examples of the lower level semantics are Business Process Patterns – formal and explicit descriptions of the generalized designs – best practices for business in a given application domain [22]. A Domain Specific Model (DSM) for business logic of information systems is proposed, based on the analysis of the modeling concepts of visual behavioral modeling languages [23]. DSM is considered as abstract business logic model (called Amabulo meta-model), combining process, state and structural perspectives.

B. Runtime models

Typically, runtime models are considered as assets which are used to monitor and verify particular aspects of the runtime behavior of the information system [24][25]. It is foreseen that the runtime models will be used by the agents responsible for managing the runtime environment, and for adapting and evolving the software during runtime. Hence, the models are considered as interfaces between running systems and change agents, where a change agent can be a human developer or a software agent [5].

Research on the runtime models is still in its infancy; however, the foreseen opportunities are beyond significant. France and Rumpe [5] predicted the possibilities of system users to observe and understand system behavior, adaptation agents to detect the need for adaptations and perform these, change agents to handle errors and introduce new features, all by using runtime models. It is obvious that these assumptions are inherited from the theory of adaptive, self-managed systems which specifically deal with change processes [26].

In the above described context, runtime models are expected to evolve the current practices of MDE, from design, implementation and verification stages of development of the EIS, to their actual execution. The evolution of models and associated approaches to a change management is considered only as a first step, towards the vision of EIS's as shells which are actually executing models, which embed all application aspects, including data and information, business logics and user interface.

With the emergence of semantic technologies, some initial works on developing ontology-driven systems, using formal models, expressed in RDF/OWL, have been carried out.

C. Ontology-driven systems

Today, ontologies are increasingly used to facilitate a process of software design. However, in great most of the cases, their use is related to maintaining different schemas (with increased expressiveness when comparing to conventional semi-structured data approach, such as XML) and to providing the formal foundation to conventional MDE environments, thus enabling their verification.

Although the term of ontology-driven information system, as a system that make use of formally defined ontologies, was coined by Guarino in 1998 [27], the use of conceptual schemas to represent the knowledge about specific application domain was proposed as early as in the seventies [28].

The common use of ontology for information systems is related to facilitating cross-domain explanation and understanding of invariants of one domain. There are opinions that it should not be confused with the different conceptual schemas (e.g. ER, UML, OMT) used in systems' modeling, as latter involve specification of a meaning of these invariants in the different dimensions [29].

Here, we would like to highlight the difference between ontology-based and ontology-driven software. In the former case, even though that sometimes ontologies play the central role, a large amount of business logic is still implicitly contained in a source code. The latter case fully corresponds to runtime models paradigm, where all data structures and business logic are formally described in ontology and then interpreted by the shell software at runtime. We use the notion of "shell software" because it can be then considered as a model execution platform.

Currently, there are only few works that use the ontologies as runtime application models. When considered as data schemas, ontologies can be used to drive software which would enable managing its individuals. Such software is then no more than a tool for ontology browsing and concept instantiating. Ontology Based Information System (OBIS) [30] is an example of such an approach.

OntoWebber System Architecture [31] enabled web portal management, where ontological framework is used to integrate and semantically align different data sources in order to generate a web portal. To a minor extent, it also addressed business logic, by enabling formal description and correspondent on-demand operation of few business rules, related to personalization and web site maintenance.

One of the proposed approaches to ontology-driven software [32] considers the interaction flow as a key artifact of runtime ontology. This flow consists of the actions of registration of the specific event occurrence, its categorization, identification of the situation that matches the categorized event and consequent execution of one or more tasks (including interpretation of a business domain model to generate recommendation for some of these tasks). In fact, such conceptualization can be used to formalize business rules and is thus useful for further development of ontology-driven software paradigm.

1) Ontology Scaffolding

Ontology scaffolding is an approach in which a basic application, with so-called CRUD (Create, Read, Update and Delete) functionality is generated in design or run time, based on a specified model.

Scaffolding approach became popular with the development of MVC (Model-View-Controller) frameworks and is typically related to using database schema to create scaffolds.

D. NoSQL databases

One of the critical objectives of the design of the future shell software that will execute runtime models is related to a full independence, relative to the agreements and sometimes, compromises on the use of the persistence layer. Currently, NoSQL databases are the best educated guess for facilitating such an approach, characterized by the two advantages over traditional relational database systems: flexibility and performance.

The flexibility of NoSQL is arising from the fact that it does not have to adhere to schema definitions, which typically correspond to data model of application which is using those. Instead, information is stored in semi-structured way, by using key-value pairs, documents, graphs or wide-column stores. The light structure of the storage formalisms contributes to the high horizontal scalability of NoSQL databases. Namely, unstructured data can be more easily stored across multiple processing nodes, because it does not follow complex data structures and it avoids join operations.

When comparing the SQL and NoSQL databases, the latter are highly preferred options for large data sets and

hierarchical data structures. SQL databases still outperform NoSQL in facilitating complex transactional applications. Finally, native support to ACID makes relational databases superior to NoSQL when reliability and consistency are considered.

III. ONTOAPP TOOL

OntoApp is ontology scaffolding tool which generates CRUD functionality in runtime, based on the specified ontology – RDF/XML file. It is a PHP web application, using RDF API for PHP [33] for ontology interpretation and Neo4JPHP API for storage. It uses Neo4J graph database [34] for storing instances. Thus, it is persistence neutral in the sense that it does not adhere to a specific structure of the database as an implementation condition.

In this paper, we refer to the case of ontology-driven project management application. The application is implemented by using OntoApp with the simple project management ontology. The UML representation of the portion of project.owl ontology is illustrated on Figure 2.

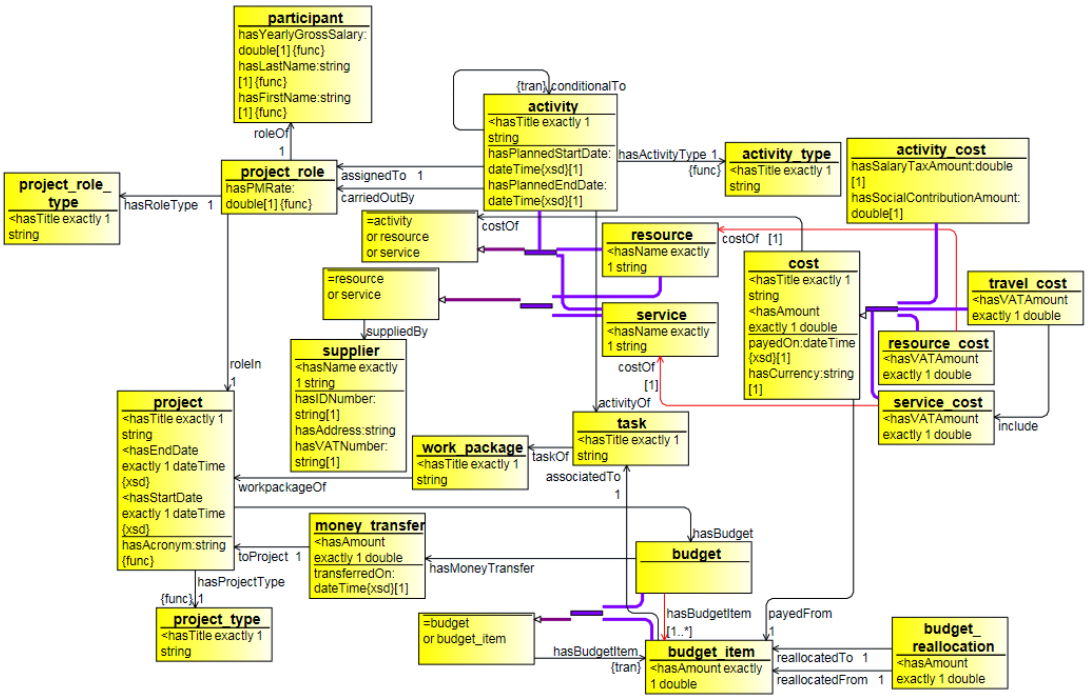


Fig.2 UML Representation of the project.owl ontology

The different views to the different concepts of example ontology in OntoApp are illustrated on Figure 3.

OntoApp interprets the formal model, expressed as RDF/XML ontology to generate CRUD functionality on the set of concepts, as specified in ontology. Based on the formal definition of each of the concepts, the form is being generated in a runtime and used to define an instance of the given concept, which is then stored in Neo4J database as a graph node, with a label corresponding to the name of the concept.

The data properties are defined as properties of a node, while each of the instantiated object property correspond to the relationship being established in a graph database between the specific node and another existing node in a domain of the object property.

Furthermore, the generated form implements certain validation rules which are determined based on the formal definition of the concept which instances are being created. OntoApp interprets the formal restrictions expressed as anonymous parent concepts – necessary conditions for a given concept. These formal restrictions are defined as value (owl:allValuesFrom, owl:someValuesFrom) and cardinality (owl:cardinality, owl:minCardinality, owl:maxCardinality) constraints of both data and object properties.

A. *Interoperability as an inherent property of OntoApp*

One of the key benefits of the formal runtime-model-driven applications is that they are inherently interoperable.

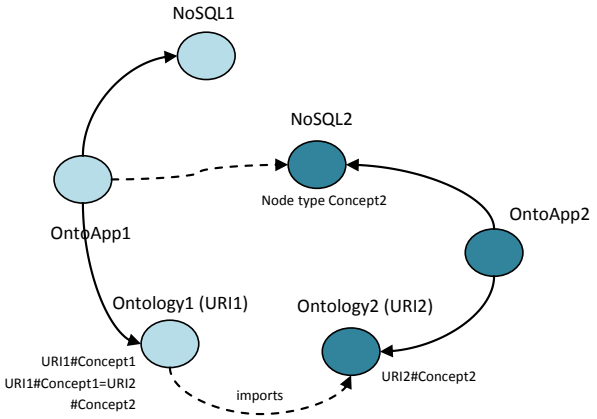


Fig.3 Illustration of the interoperability as inherent property of OntoApp

This inherent interoperability arises from the core of the approach. Namely, each OntoApp system comprises of three assets, natively distributed (see Fig.3): OntoApp, ontology and NoSQL database. OntoApp execution environments (OntoApp1, OntoApp2) are installed on the

different platforms. Each of the environments is driven by one of the respective ontologies (Ontology1, Ontology2) with specified Uniform Resource Identifiers (URI1, URI2). Then, the execution environments are using ontologies to create and manage graphs, stored in respective NoSQL databases (NoSQL1, NoSQL2).

When Ontology1, driving OntoApp1 environment imports Ontology2, the concepts of the latter can be instantiated (nodes created and managed) in NoSQL2, by using OntoApp1. This is possible only for concepts of Ontology2, which are annotated with the connection strings, corresponding to the location and authentication of NoSQL2 database; and defined access rights.

When logical equivalence relationship is established between the different concepts in source (Ontology1) and imported ontology, then OntoApp1 can be enabled with a centralized access to the distributed repository of information objects, thus enabling for example, integrated reporting, bulk processing, etc.

Finally, using other logical relationships to connect the different concepts from the different ontologies enables the construction and maintenance of the federated objects, whose different attributes are stored across multiple NoSQL repositories.

B. *Future Development*

Currently, OntoApp tool development aims at investigating possibilities to: 1) further customization of CRUD functionality, in terms of more closely adhering to the actual specific users' needs; and 2) modeling and embedding elements of the business logic into application. In order to make this possible, the helper ontology is being developed, that will enable formal definition of the rules related to restricting the accessibility to the specific concepts, instances or sub-graphs and business rules.

Access restriction is being implemented in two ways. First, simple user administration model is being embedded, based on its formal definition in the helper ontology. It defines user instances (stored in helper ontology) and assignments of CRUD rights, on the specified concepts of the imported ontology – runtime model. Second, context restriction is being implemented. It enables restriction of CRUD rights based on the selected instances – enabling access to all instances (nodes) from or to which a graph can be traversed.

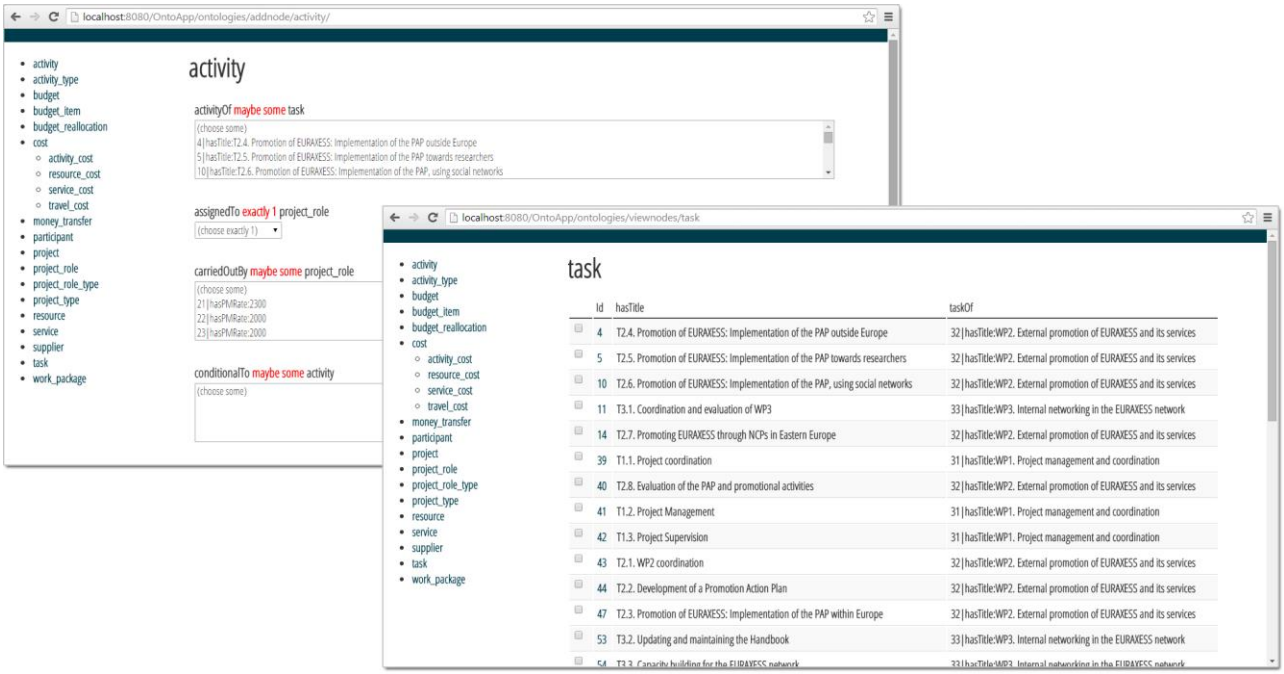


Fig.4 OntoApp

Another aspect of customization is related to enabling users to customize the layout of the app by themselves, e.g. to rearrange order of properties in add/edit forms, build the custom menus, etc.

The work on modeling and embedding elements of the business logic into app is still in initial phase and thus, it will not be considered in this paper.

IV. CONCLUSION

Current research on runtime models consider their application on the traditional EIS architecture, where the runtime models are used by the change agents. Although this work is still in initial phase, the research on the runtime models with full functional coverage and respective execution environments is expected to gain the attention of the research community, due to widening gap between the traditional approaches to EIS design and development and rapidly growing diversity of identifiable processing platforms.

Besides the significant potential impact, related to dramatically improved flexibility and facilitation of so-called self-* properties of EIS, e.g. self-management, self-optimization, etc., the use of runtime models and their execution environments is also expected to simplify and accelerate innovation, since they facilitate rapid verification and validation of new concepts and ideas.

In its current shape, OntoApp is a simple and yet functional tool for managing information in a specific domain, where this domain is formally described by using the ontology. OntoApp is inherently interoperable, since it

relies exclusively on the semantic model, namely the ontology, which concepts can be made logically correspondent to the concepts of the other domain ontologies.

It is expected that further development in customization only would contribute to its usefulness beyond the limits of the academic exercise. Further research and implementation of the other different aspects, such as business rules and reporting is expected to bring the clear evidence on the concept of model execution environments, as a key feature of the future NG EIS.

ACKNOWLEDGMENT

The work presented in this paper was supported by the Ministry of Education and Science of the Republic of Serbia (project No. III41017).

REFERENCES

[1] Lee, E., Cyber Physical Systems: Design Challenges. Technical Report No. UCB/EECS-2008-8, 2008, University of California, Berkeley.

[2] Ashton, K., That 'Internet of Things' Thing, in the real world things matter more than ideas. RFID, http://www.rfidjournal.com/articles/view?4986, 2009.

[3] FInES Future INternet Enterprise Systems - Research Roadmap 2025. 2012.

[4] Santucci, G., C. Martinez, and D. Vlad-Câlcic The Sensing Enterprise. 2012.

[5] France, R., & Rumpe, B. (2007). Model-driven Development of Complex Software: A Research Roadmap. Proceedings of Future of Software Engineering (FOSE 07) (pp. 37 - 54). Washington, DC, USA: IEEE Computer Society.

[6] Soley, R. M., Frankel, D., Mukerji, J., & Castain, E. (2001). Model Driven Architecture - The Architecture of Choice. OMG.

[7] OMG Adopted Specification ptc/03-10-04. The Meta Object. OMG.

[8] (2005). The Object Management Group - UML 2.0: Superstructure Specification. Version 2.0.

[9] QVT-Merge Group 1.8. Revised submission for MOF 2.0. OMG.

[10] Czarnecki, K., & Helsen, S. (2006). Feature-based survey of model transformation approaches. IBM Systems Journal , 45 (3), 621 - 645.

[11] McUmber, W. E., & Cheng, B. H. (2001). A general framework for formalizing UML with formal languages. Proceedings of the 23rd International Conference on Software Engineering (pp. 433-442). Washington, DC, USA: IEEE Computer Society.

[12] Abrial, J.-R., Schuman, S. A., & Meyer, B. (1980). A Specification Language. In A. M. Macnaghten, & R. M. McKeag, On the Construction of Programs. Cambridge University Press.

[13] Jackson, D. (2002). Alloy: a lightweight object modelling notation. ACM Transactions on Software Engineering and Methodology , 11 (2), 256 - 290.

[14] Jackson, D., Schechter, I., & Shlyakhter, I. (2000). Alcoa: the Alloy constraint analyzer. Proceedings of the 2000 International Conference on Software Engineering (pp. 730 - 733). Limerick: IEEE.

[15] Martin, W. 1989 Proving Program Refinements and Transformations. PhD Thesis, Oxford University

[16] Cheatham, T. E., and Wegbreit, Ben. A Laboratory for the Study of Automating Programming Proc AFIPS 1972 Spring Joint Computer Conf., 1972

[17] Bauer, F. L., Moller, B., Partsch, H., & Pepper, P. (1989). Formal program construction by transformations-computer-aided, intuition-guided programming. IEEE Transactions on Software Engineering , 15 (2), 165 - 180.

[18] Schroff, M., & France, R. B. (1997). Towards a formalization of UML class structures in Z. Proceedings of the Twenty-First Annual International Computer Software and Applications Conference (pp. 646 - 651). Washington, DC: IEEE.

[19] Dupuy, S., Ledru, Y., & Chabre-Peccoud, M. (2000). An Overview of RoZ : A Tool for Integrating UML and Z Specifications. Advanced Information Systems Engineering. Lecture Notes in Computer Science. 1789, pp. 417 - 430. Springer Berlin Heidelberg.

[20] Anastasakis, K., Bordbar, B., Georg, G., & Ray, I. (2010). On challenges of model transformation from UML to Alloy. Software & Systems Modeling , 9 (1), 69 - 86.

[21] Csertdn, G., Huszerl, G., Majzik, I., & Pap, Z. (2002). VIATRA - visual automated transformations for formal verification and validation of UML models. Proceedings of 17th IEEE International Conference on Automated Software Engineering (pp. 267 - 270). IEEE.

[22] Barros, O. (n.d.). A Novel Approach to Joint Business and Information System Design.

[23] Brückmann, T., & Gruhn, V. (n.d.). A Domain Specific Model to Support Model Driven Development of Business Logic.

[24] Bencomo, N., Blair, G., & France, R. (2007). Summary of the Workshop Models@run.time at MoDELS 2006. Models in Software Engineering, Lecture Notes in Computer Science. 4364, pp. 227 - 231. : Springer.

[25] Bencomo, N., France, R.B., Cheng, B.H.C., Asmann, U. (Eds) Models@runtime: Foundations, Applications and Roadmaps. Lecture Notes in Computer Science, Vol.8378, Springer

[26] Bradbury, J. S., Cordy, J. R., Dingel, J., & Wermelinger, M. (2004). A Survey of Self-Management in Dynamic Software Architecture Specifications. Proceedings of the International Workshop on Self-Managed Systems (pp. 28 - 33). Newport beach, CA, USA: ACM.

[27] Guarino, N. (1998). Formal Ontology and Information Systems. In N. Guarino, Formal Ontology in Information Systems (pp. 3 - 15). Amsterdam, Netherlands: IOS Press.

[28] Tsichritzis, D., & Klug, A. C. (1978). The ANSI/X3/SPARC DBMS Framework Report of the Study Group on Dabatase Management Systems. Information Systems , 3 (3), 173 - 191.

[29] Fonseca, F., & Martin, J. (2007). Learning The Differences Between Ontologies and Conceptual Schemas Through Ontology-Driven Information Systems. Journal of the Association for Information Systems , 8 (2), 129 - 142.

[30] Zviedris, M., Romane, A., Barzdins, G., & Cerans, K. (2014). Ontology-Based Information System. Semantic Technology. Lecture Notes in Computer Science (pp. 33 - 47). Seoul, South Korea: Springer International Publishing.

[31] Jin, Y., Decker, S., & Wiederhold, G. (2001). OntoWebber: Model-Driven Ontology-Based Web Site Management. Proceedings of SWWS'01, The first Semantic Web Working Symposium, (pp. 529 - 547). Stanford, CA, USA.

[32] Hammitt, L. C., & Beckert, J. (2007). Patent No. US7200563 B1. USA.

[33] Oldakowski, R., Bizer, C. 2004. RAP: RDF API for PHP

[34] Miller, J.J. 2013. Graph Database Applications and Concepts with Neo4j. In Proceedings of the Southern Association for Information Systems Conference, Atlanta, GA, USA March 23rd-24th, 2013

A Meta-metadata Ontology Based on ebRIM Specification

Igor Cverdelj-Fogaraši, Goran Sladić, Stevan Gostojić, Milan Segedinac, Branko Milosavljević
{igor.fogarasi, sladicg, gostojic, milansegedinac, mbranko}@uns.ac.rs
Faculty of Technical Sciences, University of Novi Sad

***Abstract* — This paper describes an approach for enabling unified, document type independent semantic web-based reasoning over various metadata sources in a document management system. A comprehensive, yet concise non-domain-specific metadata ontology influenced by ebXML RegRep standard is proposed to serve as a semantic basis for other domain-specific metadata ontologies to be mapped. To overcome disparities between non-domain-specific and domain-specific metadata ontologies, SWRL rules are combined with OWL knowledge base, thus enabling comprehensive semantic web-based reasoning using an extended set of OWL axioms. Dublin Core metadata ontology is used as a case study of this research.**

I. INTRODUCTION

One major limitation of existing DMSs (Document Management Systems) is the lack of domain specific services (such as domain specific browsing and retrieval of documents, life-cycle management, etc.) leading to a complicated customization of DMS for a specific domain. Document management, per se, does not necessarily have to make use of semantic web technologies [1]. Semantic web technologies, on the other hand, can be used as powerful means for implementing semantically supported DMS. To achieve a complete semantic document management, a couple of very important things such as: document management (an abstract metadata and document model), document life-cycle management, business process management, etc. have to be implemented on a semantic level. In this paper we limit ourselves exclusively to metadata, which represents only one of the mentioned aspects of semantic DMS, while the other aspects of semantically supported document management will be the subject of our further work. Correcting the aforementioned DMS weakness is our main guideline in carrying out this research.

Looking at records stored within a DMS, various domain documents can often be found. Therefore, different metadata standards, and metadata element sets consequently, are generally used to identify document content. Sometimes even documents from the same domain, issued by different agencies, are identified with different metadata element sets. Additionally, more often than not, attributes from different metadata element sets might actually be referring to the very same thing. All these documents together with their matching metadata represent a large heap of not too useful information, in terms of semantic reasoning.

Our solution for DMS deficiencies is to introduce semantics into DMS by modeling documents, business processes and metadata using semantic web technologies [1]. This model should consist of two layers: an abstract layer which models abstract documents, business processes and metadata and a concrete, non-abstract layer which models domain specific documents, business processes and metadata. Using these models enables domain specific customization of DMS, providing a wide range of domain specific services [2].

In this paper we focus on metadata model for describing documents stored in a DMS. In order for our DMS to leverage all of the available information to the maximum extent, we implemented a non-domain-specific metadata OWL ontology [3], based on ebRIM (ebXML Registry Information Model) specification. This ontology should serve as a semantic basis for the other metadata ontologies. More specifically, this non-domain-specific ontology can be considered meta-metadata ontology for other domain-specific metadata ontologies to be mapped. The ontology we implemented represents a core ontology which allows unified, document type independent reasoning over heterogeneous metadata sources, thereby enabling the employment of implicitly contained, passive knowledge.

Although we decided to go for a more flexible semantic approach, the fundamental concepts of our solution are still largely based on ebRIM specification. There are several reasons for restricting our choice to ebRIM. Probably, the most relevant one is its ubiquitousness. Nowadays, there are many different domain-specific implementations of ebRIM specification widely available, which proved it to be applicable in various domains. Just to name a few: Geospatial information systems [4], Sensor web [5], Healthcare informatics [6], Document management systems [7], and the list goes on. Furthermore, many governmental bodies and industries are also prominent in ebXML (Electronic Business using eXtensible Markup Language) Registry adoption for electronic information management and dissemination [8].

Besides the above mentioned, another fact supposed to be in favor of ebXML Registry Information Model is that ebRIM specification is a major part of the ebXML RegRep [9] standard approved by OASIS; organization considered to be one of the largest standards consortiums for electronic commerce on the Web.

II. RELATED WORK

In this section we outline the background of the research, focusing mainly on studies using either ebXML RegRep or familiar ebXML standard, for solving related problems in different domains of application.

In paper [10], authors propose a solution for mapping CPP (Collaboration Protocol Profile) and CPA (Collaboration Protocol Agreement) specifications to OWL ontology with the aim of enhancing the semantics of

ebXML in E-commerce domain. The effort has been made in order to facilitate discovery and communication between business partners, without having to exchange large amounts of data. CPP and CPA specifications are a major part of ebXML CPPA (Collaborative Partner Profile Agreement) standard, commonly used in E-commerce. ebXML CPPA is another OASIS standard, beside ebXML RegRep, which will be discussed in more detail.

The authors of another paper [11], propose a solution for semantic enhancement of ebRIM in order to support a development of SCRR (Software Component Registry and Repository) system built upon the ebXML infrastructure, and according to the ebXML RegRep standard. Methodology discussed in this paper is directed towards extending the semantics of ebXML Registry Information Model by introducing a software component attribute ontology as a means for providing semantically richer description of software components within the registry.

The solution proposed in [12] is based on a pragmatic, metadata oriented approach. In this paper, authors introduce document type ontologies to facilitate modeling of structured metadata definitions within DMSs. The solution presented in this paper is based on ebXML RegRep standard. In order to enhance DMSs and enable ontology based document management, the authors of the paper suggested mapping OWL constructs to ebRIM, conforming to OWL extension profile recommendations. According to the proposed solution, an ontology based classification of documents within a DMS is demonstrated in the paper.

Methodologies described in [4, 5] are oriented towards web service discovery as well as collaboration and interaction in context of B2B negotiations. Authors of both papers agree that WSDL description of web services does not necessarily provide enough semantics for a successful B2B marketing. Oftentimes, a high-level description of service instances is also needed. Both proposed solutions [4, 5] are also based on ebRIM. In paper [13], the authors suggested a strategy for improvement of web services discovery by introducing semantics in ebXML registries. They suggested storing semantics, in a form of OWL constructs within ebXML registries. This makes the businesses more easily recognizable, thus making the collaboration between businesses more efficient. Another paper [14] related to web service discovery, also addresses the ebXML enrichment issue. In this paper authors describe how registries can be enriched in order to describe web service semantics through OWL ontologies. They propose matching OWL constructs to ebXML classification hierarchies, and describe how expressed semantics can be queried through standardized queries, which are an integral part of ebXML facility.

Another approach of semantic tagging and discovery of services, as well as other resources is covered in lecture [15]. Problems discussed in the lecture involve structure mapping: querying across heterogeneous data structures, and conceptual mapping: using knowledge from domain ontologies in order to improve the results. The authors recommended using domain ontologies for metadata enrichment. This is accomplished by tagging metadata with concepts from widely accepted domain ontologies, which allows any new domain knowledge to be leveraged instantly. The work presented in this lecture is based on OWL extension profile of ebXML RegRep standard, mentioned earlier in this section.

The work reported in [16] is headed towards identifying the potentials for ebRIM usage in healthcare informatics. The main focus of the research is achieving a semantic interoperability amongst different healthcare systems. The authors of the paper describe how ebXML registry semantic constructs can be utilized for annotation, storage, retrieval and discovery of medical archetypes. The solution they propose includes the implementation of archetype metadata ontology. They also describe a strategy for accessing archetype semantics through standardized queries and ebXML query facility. Moreover, the authors suggest how archetype data can be efficiently retrieved from medical information systems by using ebXML messaging system, a standard way of exchanging messages between organizations.

III. EBRIM SPECIFICATION

ebXML registry plays a fundamental role in the ebXML architecture. It serves as an application gateway for a repository to the outside world, governing how parties interact with the repository. It provides a means for sharing of relevant company information in a form of business semantics, to relevant parties in a highly controlled manner. The ebXML registry can be considered an interface, independent of the underlying network protocol stack, for accessing and discovering shared business semantics [17]. The main purpose of ebXML registry is enabling business process integration between the interested parties.

ebXML RegRep is an open specification approved as an OASIS standard [18] for metadata and content management software. It is capable of managing diverse content such as documents, images, services, devices, assets, schemas, WSDL, ontologies, records [19]; which, among other aforementioned things is the reason why ebXML is relevant, and thus so important for document management systems. On the other hand, ebRIM specification is a substantial part of ebXML RegRep standard. As such, it provides a high-level blueprint for metadata in the ebXML registry [20]. Its elements do not represent repository content, but the content metadata. They provide a definition of metadata type and their relationships, at a higher conceptual level.

There are two integral parts of ebXML RegRep standard. The first one is a registry, and the other one is a repository. The repository stores digital content, while the registry stores corresponding metadata. Accordingly, there are two major types of resources: Repository Item, represents an object stored in a repository, and Registry Object, a metadata used by a registry to classify and manage repository items. The ebXML Registry Information Model defines classes and their relationships used for Registry Object metadata representation.

Since our approach is metadata oriented, in this paper we will focus on Registry Object as a key resource representing metadata. Repository Item will be introduced later in the next section in order to establish a connection between a document (Repository Item) and its metadata (Registry Object) in ebRIM ontology.

The ebRIM model is composed of several sub-models: Core, Association, Classification, Provenance, Service, Event and Cooperating Registries information model.

Although each one of them plays an important role in a description of ebXML registry; Core, Association, Classification and Provenance information models are of a special importance in a definition of ebRIM metadata model. Therefore, they will be described and taken into consideration in more detail in the next section which deals with the implementation of its ontological counterpart.

Beside RegistryObject class, being crucial for modeling metadata, some of the other, not less significant classes include: *ExternalIdentifier* class enabling advanced methods for registry object identification; *ExternalLink* class providing a mechanism for linking registry objects to arbitrary content; *ClassificationScheme*, *ClassificationNode* and *Classification* classes introducing domain-specific taxonomies between registry objects; *RegistryPackage* class allowing aggregation of logically related registry objects; *Slot* class providing an unrestricted extensibility of registry object attributes.

In addition to the above mentioned classes, there are several equally important classes such as: *User* (super class: *Person*), *Organization*, *EmailAddress*, *TelephoneNumber*, *PostalAddress* describing parties responsible for creating, publishing or RegistryObject maintenance. As well as other Core, Association, Classification and Provenance member classes not mentioned above.

IV. EBRIM ONTOLOGY

ebXML RegRep specification is reported to be designed for extensibility. However, although there is a couple of very useful extension profile specifications; neither one of them suggests a semantic approach, except for the ebXML RegRep OWL Profile. What we found as problem with OWL extension profile is that although it does suggest an approach of mapping OWL to ebRIM in order to improve the capabilities of ebXML RegRep, it does not propose mapping in the opposite direction. Which we did in order to overcome the main problem mentioned in the introduction (Section I) of this paper.

Gruber defined ontology as "a formal specification of a shared conceptualization" [21]. Conforming to the proposed ontology definition, we implemented ebRIM ontology as a formal specification of ebXML Registry Information Model.

The ebRIM ontology represents a non-domain-specific metadata ontology. Its main purpose is to serve as meta-metadata ontology for other domain-specific metadata ontologies like: Dublin Core, ISO 82045, ISO 19115 and VRA Core. The ebRIM ontology can be considered a semantic basis for mapping other domain-specific metadata ontologies. The solution we propose in this section is implemented with the aim of enabling unified, document type independent SPARQL [22] based reasoning over heterogeneous metadata sources; at the same time leveraging the contained knowledge, as efficiently as possible.

A. Implementation of ebRIM Ontology

In this subsection we suggest one approach, based on ebRIM specification, for solving the main problem of this research.

According to previously mentioned ebRIM specification [23] we implemented unique metadata ontology in order to achieve semantic reasoning over various types of documents and their metadata. In this section the implementation of ebRIM ontology will be discussed in more detail, focusing mainly on the implementation strategies we used for creating the ontology.

The ebRIM ontology is implemented across several key layers, with each layer representing its fundamental concepts: core, classification, association and provenance. More specifically, ontology consists of four sub-ontologies: (1) *Core*, (2) *Classification*, (3) *Association*, and (4) *Provenance information model* ontology. (1) *Core information model* ontology is the main sub-ontology which defines core metadata classes, including the common base classes and relevant properties. (2) *Classification information model* sub-ontology provides a semantic means for the introduction of taxonomies amongst the aforementioned resources. On the other hand, (3) *Association information model* sub-ontology defines elements (classes and relevant properties) which in conjunction with Classification sub-ontology enable a semantically richer way of expressing many-to-many relationships between resources. Whereas, (4) *Provenance information model* sub-ontology represents a typical top-level ontology, made up of classes which enable description of provenance (source information) for the concepts proposed in Core ontology.

Figure 1 shows the ebRIM ontology class hierarchy, focusing on its affiliation to the key sub-ontologies by highlighting the graph nodes. Highlights, in a form of colored circles on the upper edge of each graph node, indicate the affiliation of the class to the corresponding sub-ontology. As previously mentioned, ebRIM ontology consists of four sub-ontologies: Core, Classification, Association and Provenance information model ontology, with each sub-ontology representing a semantic counterpart of a matching model defined by ebRIM specification; which means, the proposed ontology, notwithstanding minor changes, is largely based and highly influenced upon the modeling patterns of ebXML RIM.

Core information model ontology classes, marked with a bright shade of red (Fig. 1), are the following (given in alphabetical order): *ExternalIdentifier*, *ExternalLink*, *ExtrinsicObject*, *Identifiable*, *ObjectRef*, *RegistryObject*, *RegistryPackage*, *RepositoryItem*, *Slot* and *VersionInfo*. The role of each class is elaborated below:

- *ExternalIdentifier* class facilitates RegistryObject identification with external identity information. ExternalIdentifier instance has an identificationScheme object property, referencing the ClassificationScheme instance to identify external identifier type.
- *ExternalLink* class is used to associate a registry object with arbitrary content residing either in or outside the registry.

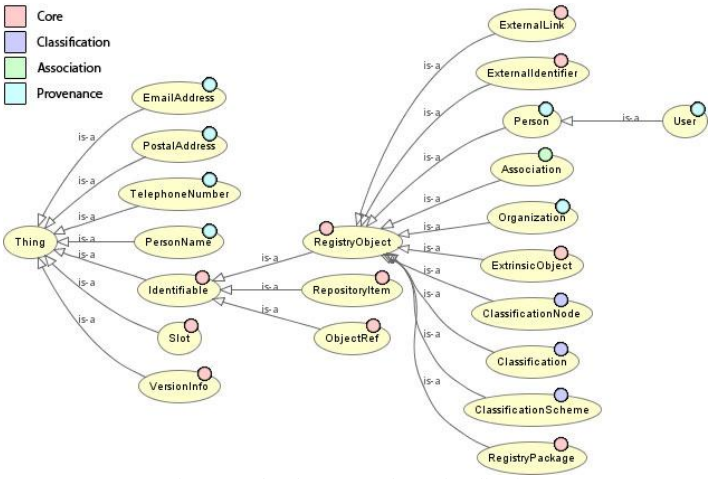


Figure 1. The ebRIM ontology visualization

- *ExtrinsicObject* class provides another strategy for ebRIM extensibility. ExtrinsicObject instances represent a repository item metadata of non-intrinsically known type, i.e. whose type is not known to the registry. It is a primary metadata class for a RepositoryItem.
- *Identifiable* class is a super class for the most classes within the ebRIM. Information model classes, whose instances require a unique identifier, are derived from the Identifiable class.
- *ObjectRef* class models a reference to a registry object. An ObjectRef instance is used to reference a RegistryObject instance.
- *RegistryObject* class serves as a common base type for all metadata elements in ebRIM ontology. Additionally, it can be considered an extensible metadata container, since it is responsible for holding the metadata.
- *RegistryPackage* class allows semantically related RegistryObject instances to be grouped together in a RegistryPackage.
- *RepositoryItem* class represents a repository item, the owner of metadata. RepositoryItem instances can be considered as an abstract view of a document.
- *Slot* class provides a means for extensibility within the ebRIM, whereas Slot instances serve as extensible attributes for registry objects. They enable a dynamic way of adding arbitrary attributes to registry entries, making the information model flexible in modeling various metadata standards.
- *VersionInfo* class represents information about the specific version of RegistryObject.

Above mentioned Core sub-ontology classes represent a subset of classes proposed by Core information model specification, with the addition of RepositoryItem class, which represents the owner of metadata, i.e. an abstract view of a document. Although this class originally does not belong to ebRIM, the introduction of RepositoryItem plays an important role in the establishment of a connection with metadata, semantically modeled by RegistryObject class. This addition is aimed at facilitating the extensibility of meta-metadata ontology, as it allows an easy way of semantic linking with other, widely used document-based ontology models, which enables even more comprehensive semantic reasoning.

Classification information model ontology classes, marked with a bright shade of magenta (Fig. 1), are the following:

- *Classification* instance explicitly classifies a RegistryObject instance, by referencing a ClassificationNode instance, defined within a ClassificationScheme.
- *ClassificationNode* instance enables refinement of ClassificationScheme tree-like structures. Taxonomy trees are constructed by nesting ClassificationNode instances underneath a ClassificationScheme instance.
- *ClassificationScheme* class equips ebRIM with a means for associating hierarchical information in a form of domain-specific taxonomies to RegistryObjects. ClassificationScheme instances represent tree-like structures (taxonomy trees) for classification of RegistryObject instances.

Association information model consists of only one class, marked with a bright shade of green (Fig. 1), whose role is to provide a strategy for associating two RegistryObject instances. An Association instance represents a many-to-many relationship between two RegistryObject instances. Association instance has an associationType object property, pointing to the ClassificationScheme instance, used to identify its type linking it to a domain-specific taxonomy.

Provenance information model ontology classes, marked with a bright shade of blue (Fig. 1), are the following:

- *Person* instance represents a person or human being.
- *User* instances represent a user known to the registry.
- *Organization* instances provide general information about organizations, where each Organization instance may have a reference to its parent.
- *PostalAddress* class defines data type properties of a postal address.
- *EmailAddress* class defines data type properties of an email address.
- *TelephoneNumber* class defines data type properties of a telephone number.

In the next section, a case study of mapping Dublin Core metadata element set to ebRIM meta-metadata ontology is further elaborated.

V. CASE STUDY: MAPPING DUBLIN CORE TO EBRIM

After the implementation of ebRIM ontology, another step is to be taken in order for a DMS to fully leverage the benefits of using a semantic approach for metadata modeling. That step includes mapping domain-specific metadata ontologies to ebRIM ontology. In the case an appropriate ontological representation of a domain-specific metadata does not exist, mapping of element set to a domain-specific ontology precedes ebRIM ontology mapping. An example of such a case would be mapping of Akoma Ntoso [24], an XML schema for description of parliamentary, legislative and judiciary documents, to ebRIM metadata ontology. Creating an OWL ontology based on Akoma Ntoso metadata element set described in XML schema, in this case, precedes mapping between these two ontologies, Akoma Ntoso and ebRIM.

As a case study of our research we decided to use Dublin Core, a domain-specific metadata element set [25]. The reason we chose Dublin Core over a number of other domain-specific metadata alternatives is that it is equally used to describe web as well as physical resources, which is why it is so widely employed as a metadata standard.

The most common ontological representation of Dublin Core metadata element set available online is protégé-dc [26] proposed by the authors from Stanford University. Since it is mainly intended to provide the annotational information for other ontologies, the protégé-dc implementation suggests annotation properties for modeling metadata elements. However, due to the restrictions of annotation properties, such implementation does not provide enough means for efficient semantic reasoning. According to OWL Reference document [27], annotation properties provide no semantics in OWL DL, and therefore are completely ignored by the reasoner.

Considering the limitations of annotation properties in terms of the semantic reasoning, we did not find protégé-dc ontology to meet our requirements. For this reason, we implemented (Fig. 2) a unique Dublin Core metadata element set ontology using OWL constructs which enable efficient semantic reasoning.

As shown in Figure 2, there are two major classes on the same level of the hierarchy: *Metadata* and *Field*. Metadata class allows modeling Dublin Core metadata element set. Whereas the elements are modeled as Fields of which each Metadata instance consists. Class Field provides a super class for other classes representing the 15 core elements defined in Dublin Core specification: *Contributor*, *Coverage*, *Creator*, *Date*, *Description*, *Format*, *Identifier*, *Language*, *Publisher*, *Relation*, *Rights*, *Source*, *Subject*, *Title* and *Type*.

The relation between Metadata and Field individuals is modeled as an object property *hasField*. On the other hand, actual metadata content of each Field individual is linked with a data type property *textContent*.

Linking domain-specific metadata ontologies to ebRIM ontology is the final step in enabling semantic web-based reasoning over various metadata sources. In order to overcome disparities between Dublin Core and ebRIM ontologies, and define exact semantic rules for mapping different concepts, we introduced SWRL rules.

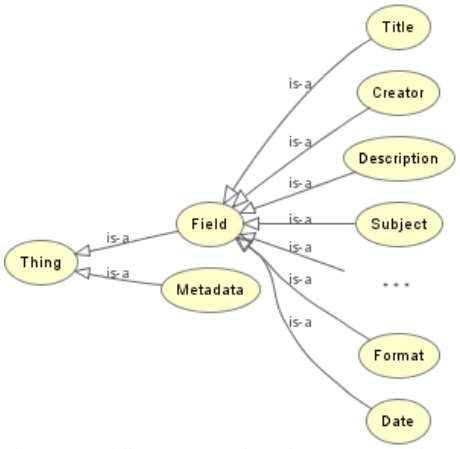


Figure 2. Dublin Core metadata element set ontology

Due to their expressiveness, we used SWRL (Semantic Web Rule Language) [28] rules in a form of Horn clauses, as a means for mapping concepts over various metadata sources. SWRL rules are combined with OWL knowledge base, thus enabling comprehensive semantic web-based reasoning using an extended set of OWL axioms.

Listing 1 shows SWRL rules used for mapping typical concepts between these two ontologies, where each rule represents a set of similar rules relevant for mapping a corresponding set of Dublin Core classes. The first set of rules applies to: *Contributor*, *Creator*, *Publisher*, *Subject*, *Date*, *Type*, *Format*, *Language*, *Coverage* and *Rights*; the second: *Title* and *Description*, the third: *Identifier*, whereas the last one provides a rule for mapping: *Source* and *Relation* Dublin Core ontology classes to ebRIM.

Mapping of other domain-specific metadata ontologies to ebRIM ontology may be achieved in a similar manner, following the very same concept defined in this section.

VI. CONCLUSION

In this paper we presented an approach for enabling unified, document type independent semantic web-based reasoning over various metadata sources in a DMS. The proposed solution is influenced by ebXML RegRep standard, rested on semantic web technologies. We implemented a non-domain-specific metadata ontology based on ebRIM specification. The main purpose of ebRIM ontology is to serve as meta-metadata ontology for other domain-specific ontologies. It is implemented with the aim of providing a comprehensive, yet rather concise semantic basis for mapping other domain-specific ontologies. As an example of domain-specific metadata ontology, without loss of generality, we used Dublin Core metadata element set. To overcome disparities between Dublin Core and ebRIM ontologies and define exact inferential rules for mapping different concepts, we used SWRL rules as a means for mapping concepts over various metadata sources. Due to its comprehensiveness and extensibility, other domain-specific metadata ontologies can be mapped to ebRIM ontology using the same concept, which proves ebRIM to be an ontology capable of dealing with a wide range of domain-specific metadata. At this point, ebRIM ontology does not support all aspects of ebRIM specification, which might be restricting in terms of modeling mostly document metadata. Information models such as: Service, Event and Cooperating registries, were not included in the current

```
1° RegistryObject(?regObj) ∧ Slot(?slot) ∧ Metadata(?metaData) ∧ Subject(?field) ∧ hasField(?metaData, ?field) ∧
   isMappedTo(?metaData, ?regObj) ∧ slots(?regObj, ?slot) ∧ textContent(?field, ?text) → slotName(?slot, "Subject") ∧
   slotValue(?slot, ?text)
2° RegistryObject(?regObj) ∧ Metadata(?metaData) ∧ Title(?field) ∧ hasField(?metaData, ?field) ∧ isMappedTo(?metaData,
   ?regObj) ∧ textContent(?field, ?text) → name(?regObj, ?text)
3° RegistryObject(?regObj) ∧ Metadata(?metaData) ∧ Identifier(?field) ∧ hasField(?metaData, ?field) ∧
   isMappedTo(?metaData, ?regObj) ∧ textContent(?field, ?text) → lid(?regObj, ?text)
4° ExternalLink(?extLink) ∧ RegistryObject(?regObj) ∧ Metadata(?metaData) ∧ Source(?field) ∧ externallyLinks(?regObj,
   ?assoc) ∧ isExternallyLinkedBy(?extLink, ?assoc) ∧ hasField(?metaData, ?field) ∧ isMappedTo(?metaData, ?regObj) ∧
   textContent(?field, ?text) → externalURI(?extLink, ?text) ∧ linkType(?extLink, "Source")
```

Listing 1. SWRL rule-based mapping between ontologies

version, since they were considered of less importance for document metadata modeling. However, they might be an important addition in extending ebRIM ontology towards metadata related to resources other than documents. We also plan some enhancements in the field of Provenance Information model ontology, which can be semantically enriched by incorporation of standardized top-level ontologies such as: FOAF, ORG, PROV-O and other ontologies for semantic handling of document related provenance information.

REFERENCES

[1] Berners-Lee, T., Hendler, J. and Lassila, O., The Semantic Web, Scientific American, 2008.

[2] Gostojic, S., Sladic, G., Milosavljevic, B., Zaric, M. and Konjovic, Z., "Semantic Driven Document and Workflow Managament", International Conference on Applied Internet and Information Technologies (AIIT), 2014.

[3] World Wide Web Consortium (W3C), (2012). OWL 2 Web Ontology Language Document Overview [online] Available at: http://www.w3.org/TR/owl2-overview/ [Accessed 20 Nov. 2014].

[4] Chen, X., Zhu, X., Zhang, X. and Du, D. "Geospatial data and services semantic share based on ebXML registry", Second International Conference on Space Information Technology, 2007.

[5] Chen, N., Di, L., Yu, G., Gong, J. and Wei, Y., "Use of ebRIM-based CSW with sensor observation services for registry and discovery of remote-sensing observations", Computers & Geosciences, 2009.

[6] Dogac, A., Laleci, G., Kabak, Y., et al., "Exploiting ebXML Registry Semantic Constructs for Handling Archetype Metadata in Healthcare Informatics", International Journal of Metadata Semantics and Ontologies, 2006.

[7] Bechini, A., Tomasi, A. and Viotto, J., "Enabling ontology-based document classification and management in ebXML registries", Proceedings of the 2008 ACM Symposium on Applied Computing (SAC), Fortaleza, Ceara, Brazil, 2008.

[8] OASIS ebXML Registry Webinar, (2005). [online] Available at: https://www.oasis-open.org/presentations/registry_webinar_05.ppt [Accessed 20 Nov. 2014].

[9] OASIS ebXML RegRep Version 4.0. Overview Document. (2012). [online] Available at: http://docs.oasis-open.org/regrep/regrep-core/v4.0/regrep-core-overview-v4.0.pdf [Accessed 20 Nov. 2014].

[10] Arsic, B., Đokic, M. and Stefanovic, N., "Mapping ebXML standards to ontology", International Conference on Information Society and Technology (ICIST), 2014.

[11] Song, D., Liu, W., He, Y. and He, K., "Ontology Application in Software Component Registry to Achieve Semantic Interoperability", Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC), 2005.

[12] Bechini, A., Tomasi, A. and Viotto, J., "Enabling ontology-based document classification and management in ebXML registries", Proceedings of the ACM Symposium on Applied Computing (SAC), 2008.

[13] Bahaj, M. and Baroudi, S., "Integrating Ontologies Into ebXML Registries for Efficient Service Discovery", International Journal of Applied Engineering and Technology, 2014.

[14] Dogac, A., Kabak, Y. and Laleci, G.B., "Enriching ebXML Registries with OWL Ontologies for Efficient Service Discovery", 14th International Workshop on Research Issues in Data Engineering (RIDE-WS-ECEG), Web Services for E-Commerce and E-Government Applications, Boston, MA, USA, 2004.

[15] Houbie, F. et al. "Semantic Metadata in Catalogue", Ontology and Discovery Workshop, Lecture conducted from ESRIN, 2009.

[16] Dogac, A., Laleci, G., Kabak, Y., et al., "Exploiting ebXML Registry Semantic Constructs for Handling Archetype Metadata in Healthcare Informatics", International Journal of Metadata Semantics and Ontologies, 2006.

[17] E. Chiu, ebXML Simplified: A Guide to the New Standard for Global E-Commerce. New York: John Wiley, 2002.

[18] OASIS RegRep Technical Committee [online] Available at: https://www.oasis-open.org/committees/regrep [Accessed 20 Nov. 2014].

[19] OASIS RegRep Wiki. (2014). [online] Available at: https://wiki.oasis-open.org/regrep/ [Accessed 20 Nov. 2014].

[20] Sounderpandian, J. and Sinha, T. E-business process management: Technologies and Solutions. Hershey: Idea Group Publishing, 2007.

[21] Gruber, T. R. (1992). What is an Ontology? [online] Knowledge Systems, Stanford University. Available at: http://www-ksl.stanford.edu/kst/what-is-an-ontology.html [Accessed 20 Nov. 2014].

[22] World Wide Web Consortium (W3C), (2013). SPARQL 1.1 Overview. [online] Available at: http://www.w3.org/TR/sparql11-overview/ [Accessed 20 Nov. 2014].

[23] OASIS ebXML RegRep Version 4.0. Registry Information Model (ebRIM). (2012). [online] Available at: http://docs.oasis-open.org/regrep/regrep-core/v4.0/os/regrep-core-rim-v4.0-os.pdf [Accessed 20 Nov. 2014].

[24] Akomantoso.org, (2014). Akoma Ntoso — Site. [online] Available at: http://www.akomantoso.org/ [Accessed 17 Feb. 2015].

[25] DublinCore, (2012). DCMI Metadata Terms. [online] Available at: http://dublincore.org/documents/dcmi-terms/ [Accessed 26 Nov. 2014].

[26] Protégé, (2008). Stanford DC OWL. [online] Available at: http://protege.stanford.edu/plugins/owl/dc/protege-dc.owl [Accessed 26 Nov. 2014].

[27] World Wide Web Consortium (W3C), (2009). OWL Web Ontology Language Reference. [online] Available at: http://www.w3.org/TR/owl-ref/ [Accessed 26 Nov. 2014].

[28] World Wide Web Consortium (W3C), (2004). SWRL: A Semantic Web Rule Language. [online] Available at: http://www.w3.org/Submission/SWRL/ [Accessed 26 Nov. 2014].

New Approach to Development of Supply Chain Management Information Systems through Software Factories

Nenad Stefanovic*, Danijela Milosevic*
* Faculty of Technical Sciences, Cacak, University of Kragujevac, Serbia
nenads@kg.ac.rs; danijela.milosevic@ftn.kg.ac.rs

***Abstract*— Information systems (IS) has been a critical component of enterprises for decades. However, development of information systems is usually very time-consuming, manual, error-prone and expensive. The pressure for delivery of quality enterprise information systems has been increased with more global and competitive business environment. Organizations now compete as parts of supply chains, which poses new challenges for information systems development and integration. This requires a radical and new approach to IS development. In this paper, we introduce a new methodology for development of supply chain management information systems based on the concepts of the software factories. This methodology integrates model-driven development, modular and reusable software assets, and agile development practices. In order to demonstrate applicability and effectiveness of the approach, we present the model of the supply chain intelligence software factory which offers collaborative, accelerated and automated development of supply chain business intelligence solutions.**

I. INTRODUCTION

Over the last decades, information technology recorded continuous and significant advances, both in hardware and software domains. Software industry experienced particularly rapid advances when it comes to software products, technologies and tools.

Organizations recognized the value of the software for their operations and achievement of business goals. Over the years, information systems became a critical part of any successful organization. Weather the business strategy is innovation, cost reduction, process optimization, supply chain coordination, or higher quality level, information systems are integral part of these strategies.

The global economic downturn over the last several years, brought many challenges for information technology (IT) projects. Nevertheless, organizations continue to invest significantly in software solutions and innovations, because IT is seen as a key driver towards competitive advantage and business sustainability. This poses a huge pressure on IT professionals to architect, design and implement optimal and high quality software solutions.

The current situation in software development hasn't improved significantly during the last decade. Considering the business demands and expectations, software products are generally to complex, monolithic, and defective in terms of usability, performance, reliability, etc.

In spite of these obvious problems related to software development process and products delivered, organizations still achieve significant value, which is demonstrated by the constantly increasing demand for new products and services. This means that organizations are willing to take significant risk in order to gain benefits from the IT investments.

Due to a present economic condition, organizations are starting to realign their IT strategies. One of the main challenges for IT professionals is to better align software solution with business requirements and to deliver more flexible and agile solutions cost-effectively. The key four architecture imperatives are [1]:

- Align – more than ever, alignment between business objectives and software solutions is necessary.
- Optimize – with shrinking IT budgets, the focus will be on the optimization of existing software solutions and assets.
- Externalize – the current IT paradigm shift is going toward cloud computing and Software-as-a-Service (SaaS) model.
- Consolidate – there is a pressure to do more with less, so organizations need to consolidate IT infrastructure and services in order to reduce complexity of these systems and to reduce costs.

If we compare software development with other engineering disciplines (with much higher project success rate), we can identify the following main issues [2]:

- One-off development – software products are developed independently from other similar systems and does not leverage the knowledge gained and the assets produced for those other systems
- Monolithic systems and increasing systems complexity – software systems are usually tightly-coupled and not modular, which makes them very hard to maintain, integrate or extend.
- Working at low levels of abstraction – this produces additional overhead during development and results in systems that are difficult to maintain.
- Process immaturity – this relates to software development process which is not mature enough comparing to development processes found in other industries.
- Rapidly growing demand for software systems – in contrast to slow improvements in software development process, there is an increased demand

for distributed service-oriented applications that run on multiple computing platforms.

According to the Chaos Report [3], which is published each year, shows only moderate increase in percentage of the successful software project over the past twenty years, from 16% in 1994, to 39% in 2014. When it comes to a large software projects, the statistics is even more disturbing – only 10% is considered success (delivered within budget, on time, and according to specified requirements). About 40% of projects are challenged, and 20% are canceled.

Despite of these statistics, there hasn't been significant changes in the way we develop software. Moreover, many of the recent advancement in software design process, programming languages, and development environments, are guided with the goal to provide reuse, agility, adaptability, and better management of complexity of software solutions. What is needed is a radical shift in software development process which will enable production of quality products with less risk and lower costs.

The main question is whether we can industrialize software development? Although software product have clear differences comparing to physical products, they have also some common characteristics. Software development capacity can be increased by shifting from skills, where artefacts are created almost uniquely, from scratch by small teams or individuals, to manufacturing model, where various products are assembled from reusable assets developed by different vendors, and where certain tasks are automated. In order to achieve this, the concept of software product lines and software supply chains it is needed to standardize processes, designs, packaging, and distribution. This is where the concept of Software Factories (SF) comes in.

A software factory is a software product line that configures extensible tools, processes, and content using a software factory template based on a software factory schema to automate the development and maintenance of variants of an archetypical product by adapting, assembling, and configuring framework-based components [4]. Software Factories do not use a generic approach, instead, they use custom developed domain-specific languages (DSL) to provide set of abstractions to fulfil the need of specific families of systems, such as supply chain management, e-commerce, banking, health, etc. Here models are used not only during the analysis and design, but throughout the entire life cycle.

In this paper, we present the concept of Software Factories, and introduce SF model for supply chain management. SCM information systems are one of the most complex software products since they include many organizations, support interconnected business processes, encompass distributed systems, heterogeneous platforms, and various data formats. The rest of the paper is organized as follows: first, we provide literature review and critical analysis of leading software architectural approaches. Then, we explain the methodology of Software Factories, along with SF Schema, SF Template, and DSL. We introduce the specialized SCM SF based on the concept of *Software+Services* that integrates model-driven development with service-oriented architecture (SOA). Furthermore, in order to demonstrate effectiveness and benefits of our approach, we provide overview of the concrete supply chain analytical information system developed based on the SCM Software Factory. Finally, we provide summary of the main benefits and advantages of the proposed approach.

II. BACKGROUND RESEARCH

The new approach to supply chain management means that companies must find a way to improve communication and information flow, thereby converting the traditional supply chain into an adaptive and real-time supply network. The theory is that this will allow companies to realize the holy grail of the supply chain - a holistic, responsive and flexible management of a network of supply chain resources that improve production and increase profitability [5].

This means that supply networks need to be not only cost-effective, but also to be [6]:

- Agile – Respond quickly to disruptions and unexpected changes in business environment and within the supply network. They need to be able to alter processes on demand and meet short term necessities better than other networks.
- Aligned – Interest of all supply network partners need to be aligned with the global supply network strategy. Global approach and collaborative planning and decision making are the keys to successful coordination.
- Adaptable – Supply network need to evolve over time and to adapt processes to meet other partner, key customer, and changing market needs.

The current economic crisis stresses even more the importance of coordination and information transparency in the supply network. Maybe more than ever, companies need to closely collaborate in planning and decision making, and manage business processes better than other networks in order to preserve sustainable development. SCM information systems play an increasingly critical role in the ability of organizations to reduce costs and improve responsiveness.

Although, big majority of leading organizations have implemented and are using some kind of SCM information systems, the researches show totally different end results. Some organizations experienced many benefits in terms of operations efficiency, reduced costs, better coordination with supply chain partners, and ultimately competitive advantage. However, there are many cases where SCM IS didn't provide expected results, and in some cases, organizations experienced many negative effects, such as mismatch of supply and demand, increased inventory costs, loss of market share, or decreased customer service level [7]. The analysis show that main reasons for these failures are related to inadequate IS strategy and architecture, inability to integrate different IS, and to deliver IS solutions within budget and defined timeframe, and with specified features.

In order to overcome these problems, organizations and software vendors strived to improve the way the information systems are produced and implemented. Lack of adequate enterprise architecture is one of the main reasons for the failed IS projects. Enterprise architecture is a strategic information asset base, which defines the business, the information necessary to operate the business, the technologies necessary to support the

business operations, and the transitional processes necessary for implementing new technologies in response to the changing business needs. When enterprise architectures work the way they should, they are a great resource in finding effective ways to better use IT. When they don't work well, they can be a very counterproductive and exhaust precious organizational resources.

There were many enterprise-architectural methodologies introduced during the last few decades. Today, the majority of the field use one of these four methodologies [8]:

- The Zachman Framework for Enterprise Architectures - Although self-described as a framework, is actually more accurately defined as a taxonomy.
- The Open Group Architectural Framework (TOGAF) - Although called a framework, it is actually more accurately defined as a process.
- The Federal Enterprise Architecture - Can be viewed as either an implemented enterprise architecture or a proscriptive methodology for creating an enterprise architecture.
- The Gartner Methodology - Can be best described as an enterprise architectural practice.

These approaches are fairly different from each other, both in goals and in approach. This can be a huge obstacle when choosing the methodology. On the other hand, they are somewhat complementing, and can be blended together according to particular organization needs.

An enterprise architecture can be an important asset in assisting an organization find better ways to use technology to support its critical business processes. Unfortunately, many organizations spend significant resources trying to create enterprise architectures, only to get inadequate, or even negative, value from these initiative. There are three primary reasons for such frequent expensive failures [9]. The first is an over-dependence on recursive object-oriented design and analysis (OODA) architectural methodologies. The second is the misunderstanding that creating an enterprise architecture requires developing a detailed blueprint of the entire organization or supply chain. And the third is a failure to deal with complexity.

The most valuable artifacts that are produced from the enterprise architecture are those which can be used across various problem domains. Software Factory pillars and the delivery goals of an architecture-driven process are in sync. Using standards-based deliverables, like software-factory schemas and domain-specific languages, to group and describe enterprise architecture components can take enterprise architecture to the next level. Capturing reusable artifacts with these templates gives organization a consistent way to deliver reusability [10].

Compared with typical enterprise architecture, a software factory schema deals with many other aspects of a software product family, such as requirements, design, testing, implementation, deployment, management, maintenance, etc. [2]. Although enterprise architecture implies a software product family, it does not explicitly identify one, or incorporate mechanisms to support family based development, such as a way to express how the members of the family differ from a family archetype. A software factory schema, on the other hand, targets a specific software product family and can be instantiated and customized to describe a specific family member in terms of its differences from the family archetype.

While enterprise architecture does not necessarily support automation, a software factory schema can be implemented by a software factory template to automate software development tasks. The focus of enterprise architecture is the design documentation, a software factory schema focuses on development artifacts.

Most of the critical innovation for the realization of the SF approach, are present, but with different maturity level. These innovations can be grouped into four areas [11]:

1. Systematic reuse

The main approach to reusability are software product families and lines. A software product line is a family of products designed to take advantage of their common aspects and predicted variabilities. The three main goals of a software product line are to reduce cost, improve delivery time, and improve quality [12]. Product line development consists of cooperation among three different constituencies: core asset development, product development, and management. When a set of software systems has common characteristics, they are candidates to become part of a product family or product line. A product line has a set of core assets upon which a shared family of systems is built. Core assets include shared components, infrastructure, tools, process, documentation, and shared architecture. On the other hand, self-description using metadata can be used to automate component discovery, selection, adaptation, assembly, configuration, deployment, and management. Assembly by orchestration follows the SOA approach and implements the Mediator patter and makes development by assembly much easier since software components can be developed independently, and then assembled and run later by orchestration engine.

2. Development by assembly

Critical innovations in the areas of platform independent protocols, self-description, variable encapsulation, assembly by orchestration and architecture-driven development are required to support development by assembly. XML and web service technologies play central role in assembly of software components.

3. Model-driven development (MDD)

MDD uses models to capture high level information, usually expressed informally, and to automate its implementation, either by compiling models to produce executables, or by using them to facilitate the manual development of executables. For the full realization of the MDD, a higher abstraction languages are needed, such as the domain-specific languages (DSL). DSL is textual or graphical programming language of limited expressiveness focused on a particular domain [13]. The use of DSL improves development productivity and model transformations.

4. Process frameworks

Rather than applying general development processes, we can specialize and tailor a formal process for a specific product family. This kind of customization makes sense only when it can be used more than once. When that is the case, it can be highly cost effective. Also, reusing highly focused process assets increases agility by eliminating work. Once a process framework is defined, micro

processes can be stitched together to support any work flow that the project requires.

Software Factories also promote the formation of software supply chains by partitioning software factory schemas, either vertically or horizontally, to shift responsibility to external suppliers [14]. This means that enterprise software delivery is amenable to many existing business process optimization techniques and a number of well-understood improvement practices can be readily introduced.

When it comes to practical usability of Software Factories, there are already examples of successful application in various industries such as manufacturing, financial, software, etc. [15, 16].

By introducing a software factory concept to enterprise software delivery, organizations can focus attention on the software supply chain, address inefficiencies in software delivery, and gain greater control and visibility into the delivery process.

In the next section, we introduce the supply chain software factory schema and template, and describe the supply chain software factory.

III. Model of Supply Chain BI Software Factory

A. Conceptual foundation of the Software Factory

A Software Factory defines a custom-made methodology for a specific group of systems using a graph of viewpoints [17]. Each viewpoint defines certain aspect of the life cycle for constituents of the system family, such as requirements elicitation, database, web service design, or web portal design. The factory relates reusable assets with each viewpoint, and delivers them in the context of that viewpoint to the developers of the system family, removing the need to search for appropriate assets, thus enabling validation, and supporting usage of manual and automatic guidance resources.

Software Factories automate the packaging and delivery of the reusable assets, including models and model-driven tools, other types of tools, such as wizards, templates and utilities, development processes, implementation components, such as class libraries, frameworks and services, and content assets, such as patterns, style sheets, help files, configuration files, and documentation. Since the software factory schema is a model, factories can be manipulated using tools. Larger factories can be created by combining smaller ones, and specialized factories by customizing generic ones.

The three important concepts underlying the software product lines are scope, variability, and extensibility [18].

1. Scope describes what software products can be developed using the product line assets. Scope is most often represented in the form of a capability or feature model.
2. Variability identifies the common and variable features defined in the scope. The parts implementing the common features are often incorporated in architecture frameworks. Variable capabilities are optional features and may be implemented only in some members of a product line.
3. Extensibility identifies extension points that can be used to add new features to the products based on a product line. Extensibility is used to incorporate functionality that is outside of the original scope of a product line.

The main SF component is The Software Factory schema. It is a model interpreted by developers and tools that describes software products, workflows used to produce the products, and assets used in the enactment of the workflows, for a specific family of software products in a given domain [2]. A software factory schema is a document that categorizes and summarizes the artifacts used to build and maintain a system, such as XML documents, models, configuration files, build scripts, source code files, SQL files, localization files, and so on, in an organized way, and that defines relationships between them, in order to maintain consistency among them.

On the other hand, the Software Factory template can be considered the instantiation of the Software Factory schema, the same way a model is an instance of the Metamodel. The template is basically the collection of all assets defined by the viewpoints of the Software Factory schema. These assets can be broadly divided into the following categories:

- Libraries and frameworks
- Guidance assets
- Domain-specific languages and designers,
- Feature models (capability models or solution capability models)

SF schema and template, together with the accompanying assets, need to be created and assembled for the specific domain, such as supply chain management.

B. Supply Chain Business Intelligence Software Factory Model

Today, most of the business software solutions provide effective automation of business transactions, but they do not provide efficient coordination outside organizations, along the supply chain. This usually leads to usage of diverse platforms, information systems and tools that support execution of complex business interactions in supply network. As a consequence, business productivity is decreased, and integration and coordination are very difficult to achieve.

Therefore, it is necessary to establish the bond between transactional systems, BI solutions and collaboration tools in a synchronized, flexible and secure manner. Ideally, the most optimal way would be composition of diverse applications into a single system, but in such a way that complex interactions among people and business entities can be easily plugged into structured business processes. Unfortunately, this is very hard to achieve as the existing applications are monolith and they are difficult to modularize. Application composition implies design of software solutions by assembling previously developed software components and also the functionalities related to personalization and customization which enables greater flexibility of the system.

The central elements of a Software Factory are a software factory schema and a software factory template based on the software factory schema. The software factory template configures extensible tools, processes,

and content to form a production facility for the product family [2].

SF defines specially tailored methodology for the specific software product family by using the grid of viewpoints [19]. Each viewpoint defines some of the life cycle aspects for the software product line members. These can be requirements definition, data warehouse design or defining web service interface. SF connects reusable software elements with each viewpoint and implements them in the context of the specific viewpoint.

A Software Factory is a development environment configured to support the rapid development of a specific type of application. The research should be carried out in order to design SCI software factory schema as a collections of reusable viewpoints and the definition of their relationship to each other. To build the specific SCI system, we must implement software factory schema by defining the DSLs (Domain-Specific Languages), patterns, frameworks, and tools it describes, packaging them, and making them available to product developers. It will also be a challenge to automate product development by defining the automated guidance that can be executed by the extensible development environment.

Our idea is to create a specific BI software factory for the supply chain domain, as shown in Figure 1. This SCI (Supply Chain Intelligence) factory should unify different viewpoints (platforms, process models, design methods and BI elements) into the integrated software solution. By using the grid of viewpoint we can categorize, summarize and relate different development artifacts such as models, XML documents, DW schemas, build scripts, SQL and MDX (Multidimensional Expressions) files, workflows, web services, etc. From each viewpoint, a certain aspect of the software can be built.

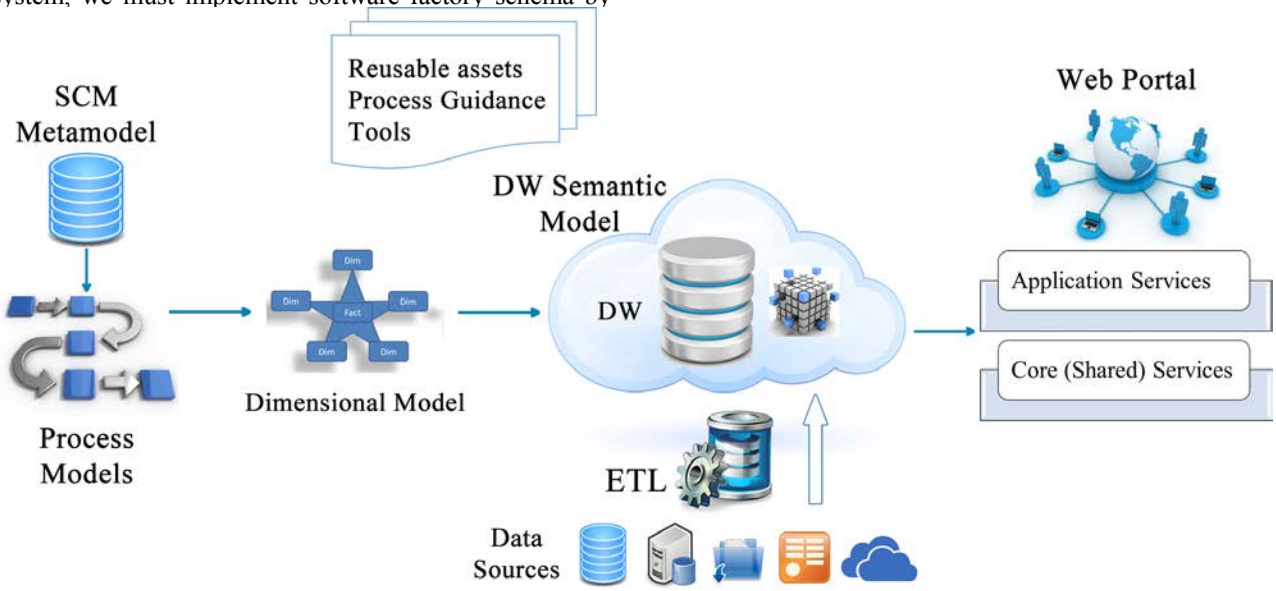


Figure 1. SCI software factory schema

In order to enable flexible supply network modelling, we have created the domain-specific SCM Metamodel [20]. The Metamodel is normalized and contains all SCOR elements such as processes, metrics, best practices, inputs and outputs. It also incorporates business logic through relationships, cardinality, and constraints. This database-centric approach enables application logic to be developed. Additionally, process standardization is the basis for development of the process and metrics repository, as one of the main elements of the SCI solution. Figure 2 shows the segment of the SCM Metamodel class diagram.

The Metamodel is extended with additional entities to support supply network modelling. In that way processes, metrics and best practices can be related to the specific node and tier in the supply network. SCOR defines processes at the three levels of detail. With this Metamodel, lower-level processes also can be modelled thus providing a more detailed view of supply chain processes and metrics.

This method offers several advantages [21]:

- Better functionality and flexibility of the model
- Metamodel contains SCM knowledge which enables domain-specific modeling.
- The usage of relational database enables integrity of data and models, data importing and exporting, as well as the option to use the standard language (SQL) for querying.
- Security and user access control.
- It is possible to design front-end web application that can serve as the interface for collaborative supply network modeling.
- Possibility to add or change both the library data (processes, metrics, best practices, etc.) and the data related to models (supply network configurations).

This approach enables modelling of any supply chain configuration and it is the basis for further modelling and data warehouse design. Based on the created supply chain model (Metamodel instance), it is possible to create data warehouse conceptual schema, with dimensions, hierarchies, measures and measure groups.

The next step is to transform DW conceptual model into the physical model and to deploy it on the OLAP (On-Line Analytical Processing) server.

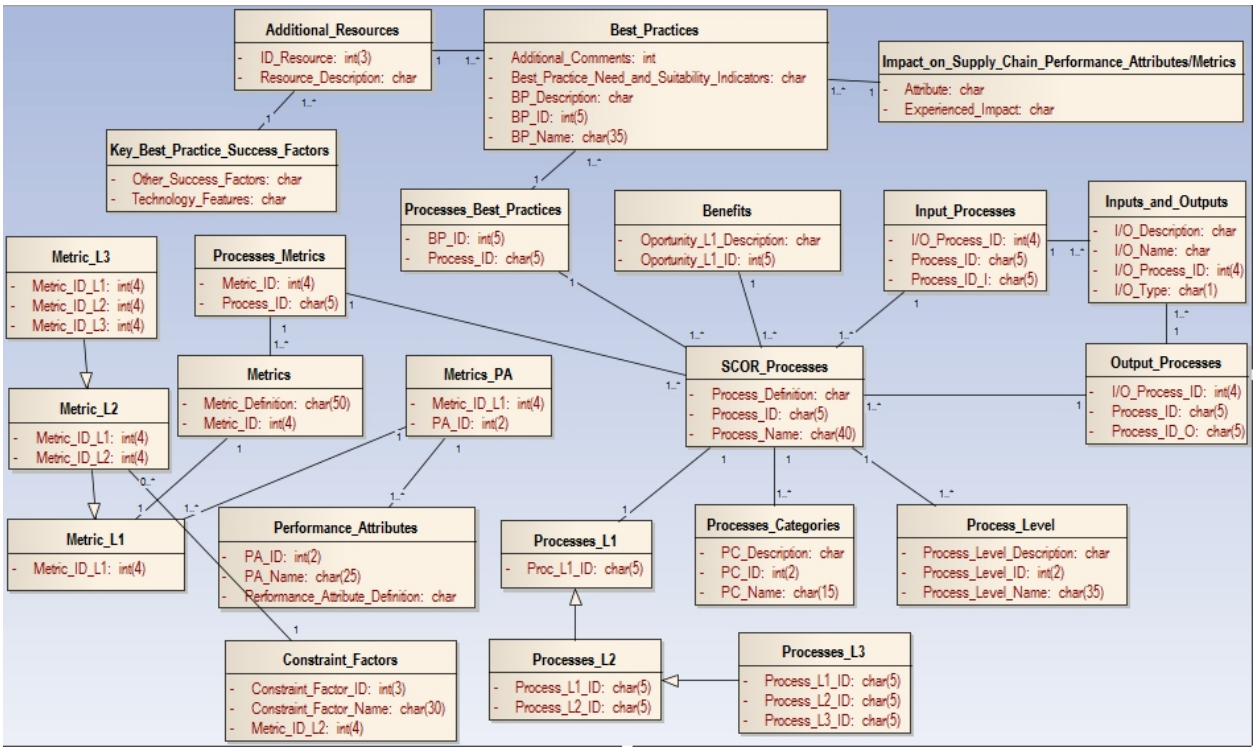


Figure 2. SCM Metamodel class diagram

For bridging the gap between the user/developer and the data sources, the BI semantic model is used [22]. A BI semantic model is constructed over many physical data sources, allowing users to issue queries against the model, using one of a variety of client tools and programming technologies. The main advantages are a simpler, more readily understood model of the data, isolation from heterogeneous backend data sources, and improved performance for summary type queries.

The final component of the SCI SF is the front-end BI portal. With its modular and extensible architecture, as well as its collaboration and analytical services, BI portal supports the design of the composite applications based on reusable templates and software assets [23].

SCI software product line architecture is open, extensible, scalable and customizable. This will enable creation of specific software supply chains where independently developed components would be easier to assemble across platform boundaries. In the supply network, each company can be the consumer and provider of information and services to other partners. This will certainly require a higher degree of standardization of packaging formats, software assets, metadata and interfaces, so that the SCI systems can be more easily discovered, connected, assembled, deployed, adapted and managed.

IV. REALIZATION OF THE SCI SOFTWARE FACTORY

In this section, we give concrete recommendations for developing a supply chain intelligence software product, using the proposed SF approach. First, we present a product development approach based on the SF schema and the template. Then, we provide a concrete example of the SCI solution with the description of the platform and reusable software assets which can be combined into specific composite applications.

A. *Product Development*

SF development, maintenance and application is a continuous process, as well as concrete product development. This process is iterative and include many activities which can be carried out in parallel and in any order, assuming that the preconditions are fulfilled. The main steps are as follows:

- Product line analysis – it supplies product specification, business case model and scope specification.
- Product line design – this includes architecture development, defining software development process, and also the process automation, where applicable.
- Product line implementation - contains implementation and process asset provisioning and packaging.

The SF schema together with the software assets (fixed and variable), such as process models, patterns, frameworks, scripts, and tools, which make up the SF template, are inputs in the product development process. Here, product specification, together with extensible development environment, customized and specific tools, are used to produce specific SCI software products.

Although product development process is usually specific to a concrete product line development, it is possible to specify an abstract product development process, which comprises problem analysis, product specification, product and collateral implementation, product evaluation, and product packaging and release, as shown in Figure 3.

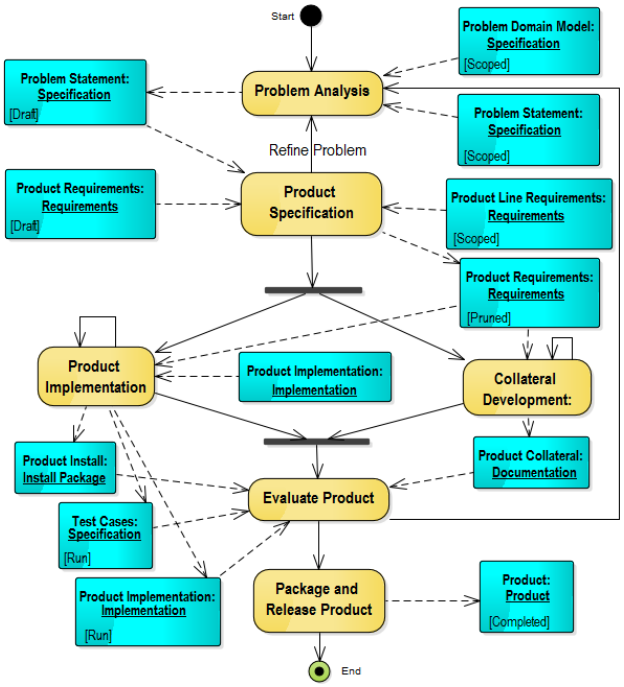


Figure 3. Product development process

This approach bring maturity in the software development process, and provides great flexibility for building similar product lines and specific products. Additionally, it speeds up significantly product development process through greater reuse of existing software assets and ensures higher product quality.

B. SCI Example

In order to realize the proposed SCI software factory model and to benefit from the composition approach, we need to view software systems granularly and per different layers of the architecture. For example, web service can be application layer asset, OLAP cube can be data layer asset, and the dashboard page can be an element of the presentation layer.

It is important to note that the set of software elements, by itself, does not necessarily make a quality software solution. We need architecture and a platform which supports application composition and enables deployment of different combinations of BI elements. Application integration and information exchange can be realized through the SOA technologies.

The containers that are provided by the platform and which hold software elements need to be of adequate type and correspond to layers of the system architecture. The system architecture is usually decomposed into three layers: presentation, application and data layers. However, three-tier architecture implies a structured business logic and data, and that all the requirements are known and defined during the system construction. By its nature, SCI applications and supply network integration are most often performed after the design and implementation of software systems. Structured business processes and traditional business applications cannot provide support for the complex interactions among supply network partners and their software systems. Because of that, we have added the additional layers to the model, namely the process layer and the collaboration layer. The architecture of the layered SCI composite application is presented in Figure 4.

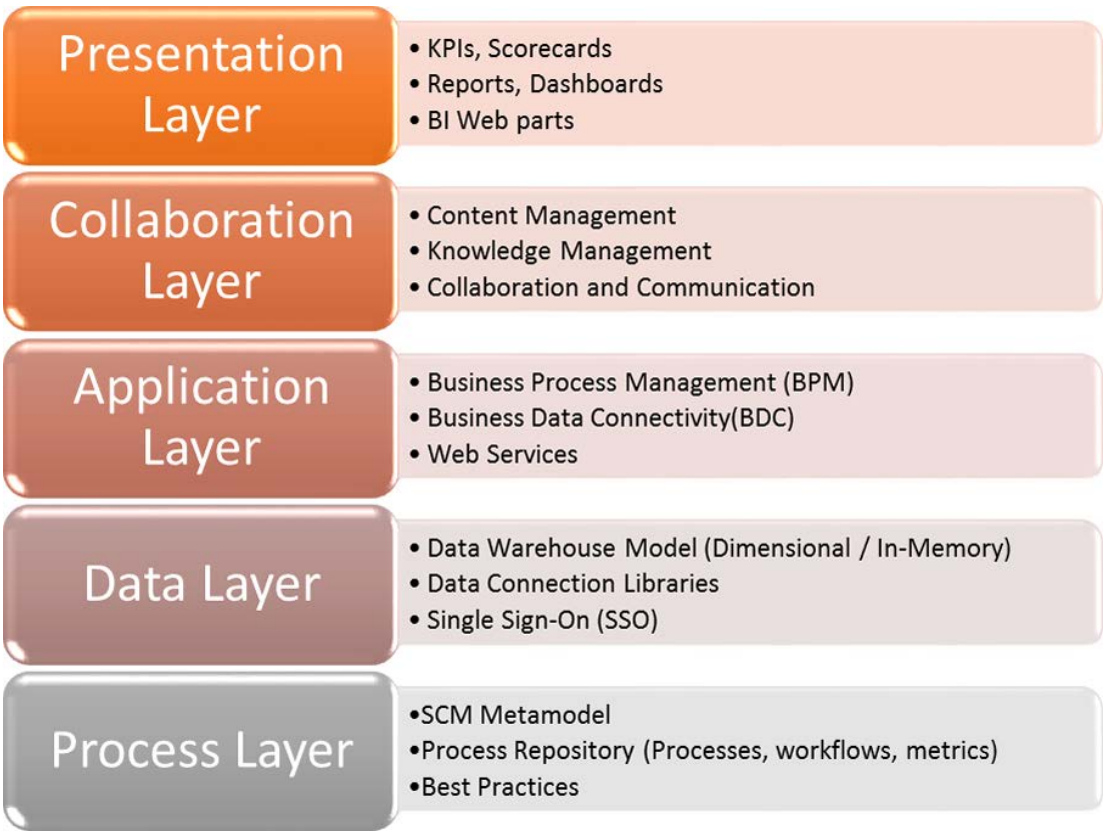


Figure 4. Layers of the SCI composite application

This architecture has advantages over classic SOA concept, as SOA typically provides flexibility only in the application layer. It offers flexibility in all the layers through the composition of reusable assets.

The design of the SCI information system consists of the following steps [24]:

1. Process design
   1) Maintenance of the process repository (process library, workflows, metrics, best practices)
   2) Process modelling via SCM Metamodel
2. Data warehouse design
   1) Dimensional modelling
   2) Design of OLAP components (cubes, KPIs, partitions, etc.)
   3) Design of ETL packages
3. Design of metadata, BI elements and packaging.
   1) Development of BI portal with specific web parts (forms, dashboards, document libraries, etc.) for particular supply network business processes. These portals can be saved as templates and reused when needed.
   2) Implementation of the BPM system (for example, BPEL-Business Process Execution Language orchestrations) for the purpose of automating business processes and for connecting portal lists and document libraries with business logic that reside on the server. These workflows can be wrapped into assemblies and then deployed on the portal.
   3) Defining data sources and BDC entities in order to integrate external systems.
   4) Composition of BI elements: reports, dashboards, spreadsheets, scorecards, etc.
4. Implementation of the BI solution
   1) Execution of the ETL packages
   2) Portal deployment to the production system.
   3) Setting the properties of BI elements
   4) Security and user management
   5) Customization and personalization of portal and its elements.

The central component is the BI portal that enables flexible and efficient design by assembly approach. It is a composite web application made of certain elements. Elements can be viewed at four distinct layers: presentation layer, collaboration layer, application layer and data layer. The portal is modular and each module can be modified and customized and also new modules can be added, thus assembling composite BI applications that fit specific user needs.

In the presentation layer, there is the following hierarchy of elements that can be combined:

- Web farm – Installation of one or more load-balanced web servers and database servers that store the basic configuration database.
- Web application – Web server site extended with the portal services and can host site collections.
- Site collection - Container for BI sites, which exists within a specific content database. A site collection contains a top-level site (supply network), with optional child sites (companies, divisions, teams or people).
- Web site - Container for child sites, pages, and content such as lists and document libraries.
- Web page - Container for web part zones, and web parts.
- Web part zone - Container for web parts.
- Web part - Components that display content on a page in modular form, and are the primary means for users to customize/personalize pages.

Composition in the collaboration layer is designed in such a way to enable more efficient and easier information creation, publication and exchange. BI portal consists of the following main elements:

- BI site - A template that can be used to create a number of BI sites with out of the box functionalities.
- Reporting module – A document library with special support for storing and managing reports.
- Dashboard - A web page assembled from different BI web parts (reports, spreadsheets, KPIs, scorecards, etc.)
- Report web part - Web part to view reports made available from external reporting server.
- Excel web part – Web part for viewing Excel sheets and graphs.
- KPI web parts – Set of web parts for used for creating, managing and displaying KPIs from different sources.

Structured business logic resides in the application layer. This can be classic ERP (Enterprise Resource Planning) application or workflow orchestration such as BPM (Business Process Management) systems. Application layer can include both transactional and analytical systems. BI portal provides option for application composition, as well as methods for consuming (integration) of external web services from other platforms. For example, business processes can be modeled using workflow activities that are deployed into the portal. Coordination of activities across the supply chain partners can be accomplished using collaboration processes that manage both the lifecycle of individual business entities (i.e. orders) and also the lifecycle of business processes (i.e. order fulfillment).

Composition in the data layer is realized through the shared platform service called Business Data Connectivity (BDC). BDC can read data from multiple types of data sources — databases, OLAP cubes, ERP systems and web services — and then return this data back into the portal through different web parts. For example, it can consume retailer's external web service which provides information about customer returns and update the dashboard for the production manager. BDC acts as a metadata repository for descriptions of business data entities and their attributes and for mappings of these entities back to data stores within the supply chain.

This multi-layered and modular architecture of the portal enables creation of various SCI applications, tailored to specific business needs. Figure 5 shows the concrete SCI analytical dashboard page composed of different modules, which are connected to different data sources or services.

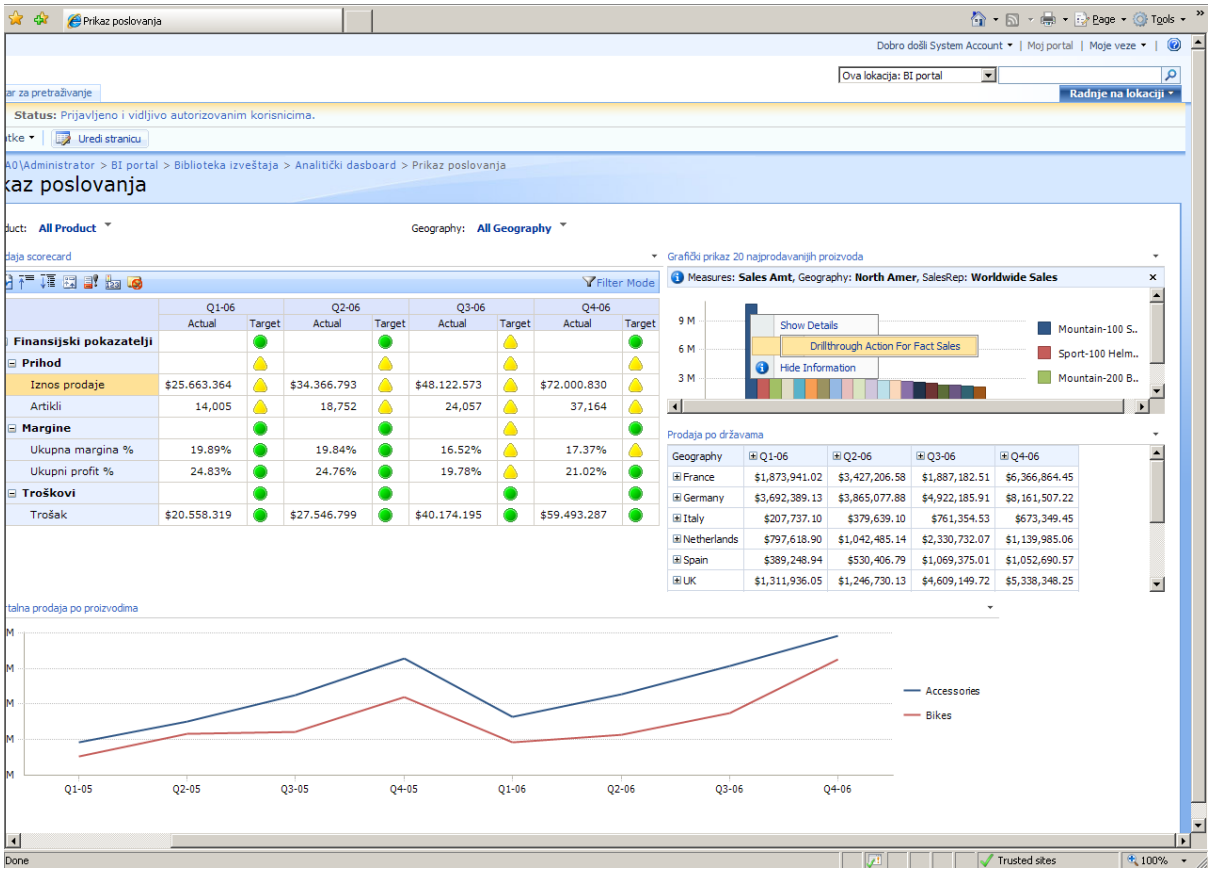


Figure 5. SCI dashboard web page

SCI software system is designed upon innovative application model and enables composition at several layers: data, services, processes and user interface.

The presented SCI software solution built based on the provided SF schema and reusable assets, provides a novel approach to design and utilization of cutting-edge web information systems for supply chain management. The main benefits can be categorized into three groups:

1. Alignment – applications can be composed in a way to accommodate requirements of all participants involved in supply chain management, monitoring, analysis and research.
2. Agility – solutions can be built, assembled, extended and deployed more rapidly and cost-effectively.
3. Adaptability – composite applications are easier to change or reconfigure when requirements change.

Although SCI SF is specific to supply chain domain and business intelligence information systems, the underlying concept of software factories is universal and can be applied to any domain. Also, certain SCI SF assets can be reused for development of different enterprise information systems, thus forming the real software supply chains.

V. CONCLUSION

Over the last few decades, organizations invested enormous amount of resources for designing and implementing business information systems. Although these information systems brought many benefits, there are still many challenges when it comes to delivering high quality information systems.

New business models, globalization, specialization and outsourcing impose more efficient and more intense collaboration among supply network partners. These trends necessitate certain changes in the way we develop supply chain information systems.

In this paper we propose the use of Software Factories for development of SCM information systems. SF can be viewed as a formation of processes, domain-specific languages, patterns, frameworks, software assets and tools, which can be used for developing set of software product variants, more rapidly, cost-effectively, and with higher level of quality.

Business intelligence software solutions are one of the most complex type of information systems, since they include many actors, processes, methods, technologies and tools. In this paper, we present the supply chain intelligence (SCI) Software Factory for designing, developing and implementing collaborative and flexible software products. The main idea behind SCI SF are process frameworks, principle of reuse, domain-specific modelling, design by composition, model-driven development, and automation. SCI SF approach will ultimately enable creation of software supply chains, where development activities, costs and risks are shared across the network of interdependent organizations.

The presented SCI information system demonstrates the usefulness and applicability of the proposed approach. The system is modular, which means that each module can be independently assembled, customized and

personalized, and existing or new modules can be combined thus assembling composite applications that suit the organization's specific needs. This modular and multi-layered architecture enables design and development of composite SCI applications which combine data, services, documents, and business process in a more creative and useful way, by assembling, connecting, and configuring the basic building blocks of functionality.

The main benefits and advantages of the SCI SF can be summarized as follows:

- Accelerated start - The Software Factory delivers an effective way for architects and developers to create a solid starting point for their application. Projects using the SF begin with a greater level of maturity than applications that are developed from scratch.
- Reduced risk - By creating a fractional implementation of a solution, which includes the most critical procedures and shared elements, architects and developers can address complex design and development challenges to expose architectural decisions and risks early in the development cycle.
- Improved quality - The Software Factory provides reusable software assets, process guidance, and tools that address common software development scenarios and challenges. SCI SF has been successfully tested for the target scenarios.
- Increased productivity - The Software Factory includes automation of certain development activities, which can be used to straightforwardly apply process guidance in more dependable and repeatable ways.
- Improved consistency - The Software Factory helps teams build multiple SCI applications and enables consistency across the design and implementation steps.
- Flexibility and customization – It is possible to model different supply chain configurations, to compose various SCI applications, as well as to customize the software factory or reusable software assets to meet the specific needs of the supply chain.

ACKNOWLEDGMENT

Research presented in this paper was supported by Ministry of Science and Technological Development of Republic of Serbia, Grant III-44010, Title: Intelligent Systems for Software Product Development and Business Support based on Models.

REFERENCES

[1] M. Walker, "Architecture in Turbulent Times," Microsoft, 2007. http://msdn.microsoft.com/en-us/library/dd547403.aspx

[2] J. Greenfield and K. Short, Software Factories: Assembling Applications with Patterns, Models, Frameworks and Tools. IN: Wiley, 2004.

[3] "Chaos Report 1994-2014," The Standish Group, 2014.

[4] J. Santos and W. Bakker, "Building Software Factories - Part 1, what are we building and why?," Microsoft, 2007. http://msdn.microsoft.com/en-us/library/bb871630.aspx

[5] SAP AG, "Adaptive Supply Chain Networks: Delivering Integrated Supply Chain Planning and Execution by Design," ASCET, vol. 7, pp. 35-37, 2007.

[6] H. Lee, "The Triple-A Supply Chain," *Harvard Business Review*, No. 10, pp. 1-10, 2004.

[7] S. Qrunfleha and M. Tarafdar, "Supply chain information systems strategy: Impacts on supply chain performance and firm performance," I*nternational Journal of Production Economics*, Vol. 147, Part B, pp. 340–350, 2004.

[8] R. Sessions, "A Comparison of the Top Four Enterprise-Architecture Methodologies," Microsoft, 2007. http://msdn.microsoft.com/en-us/library/bb466232.aspx

[9] R. Sessions, "A Better Path to Enterprise Architectures," Microsoft, 2006. http://msdn.microsoft.com/en-us/library/aa479371.aspx

[10] T. Fuller, "A Foundation for the Pillars of Software Factories," Microsoft, 2007. http://msdn.microsoft.com/en-us/library/bb245778.aspx

[11] J. Greenfield, "Problems and Innovations," Microsoft, 2004. http://msdn.microsoft.com/en-us/library/ms954817.aspx

[12] J. McGovern, S. W. Ambler, M. E. Stevens, J. Linn, V. Sharan, E. K. Jo, *A Practical Guide to Enterprise Architecture*. Prentice Hall PTR, 2003.

[13] M. Fowler, *Domain Specific Languages*. Addison-Wesley Professional, 2010.

[14] A. W. Brown, Enterprise Software Delivery- Bringing Agility and Efficiency to the Global Software Supply Chain. Addison-Wesley, 2013.

[15] V. R. Montequin, C. Alvarez, F. Ortega, J. Villanueva, "Scorecard for Improving Software Factories Effectiveness in the Financial Sector," *Procedia Technology*, Vol. 9, pp. 670–675, 2013.

[16] A. Brown, A. Lopez, L. Reyes, "Practical Experiences with Software Factory Approaches in Enterprise Software Delivery," The Sixth International Conference on Software Engineering Advances, pp. 465-470, 2011.

[17] N. Stefanovic and D. Stefanovic, "Software Factories – The New Development Paradigm," Total Quality Management & Excellence', Vol. 34, No. 3 - 4, pp. 1-6, 2006.

[18] G. Lenz and C. Wienands, *Practical Software Factories in .NET*. Apress, 2006.

[19] V. Hoogendoorn, "Software Factories 2.0 - Microsoft's larger Software Factories Initiative. Will you be a Factories 2.0 Company?," MSDN, Microsoft, 2008.

[20] D. Stefanovic and N. Stefanovic, "Methodology for modeling and analysis of supply networks," *Journal of Intelligent Manufacturing,* Vol. 19, pp. 485-503, 2008.

[21] N. Stefanovic and D. Stefanovic, "Integrated and interactive software solution for knowledge-based supply network design," *Computer Systems Science & Engineering*, CRL Publishing, Vol. 28, No. 1, pp. 5-23, 2013.

[22] R. Rad, *SQL Server 2014 Business Intelligence Development*. Birmingham, UK: Packt Publishing, 2014.

[23] N. Stefanovic, "Proactive Supply Chain Performance Management with Predictive Analytics," The Scientific World Journal, Vol. 2014. http://dx.doi.org/10.1155/2014/528917

[24] N. Stefanovic, D. Stefanovic, B. Radenkovic, "Integrated Supply Chain Intelligence through Collaborative Planning, Analytics and Monitoring," in *Integrated Supply Chain Intelligence through Collaborative Planning, Analytics and Monitoring*, I. Mahdavi, S. Mohebbi, N. Cho, Eds. IGI Global, pp. 43-92, 2011..

Prototype of a Framework for Ontology-aided semantic conflict resolution in enterprise integration

Željko Vuković*, Nikola Milanović*, Gregor Bauhoff **
* Faculty of Technical Sciences, University of Novi Sad, Serbia
** PI Informatik GmbH, Berlin, Germany

zeljkov@uns.ac.rs, mnikola@gmail.com, bauhoff@pi-informatik.de

***Abstract* — Enterprise integration carries the need for resolution of various semantic conflicts. These conflicts come in many forms and each of those may appear in a different context. Conflict detection and resolution can be made easier if a semantic description of the involved systems is available. We have developed a prototype, based on existing software - *Talend Open Studio ESB*, for a framework where a user may attach an ontology to interface elements and have those interfaces mapped automatically. We present how this prototype was tested on a scenario for which a solution was previously developed manually at Model Labs GmbH/PI Informatik GmbH, Berlin.**

I. INTRODUCTION

Developing enterprise integration solutions presents various challenges: unreliable and slow networks, heterogeneous applications, inevitable changes over time [1]. Differences between applications may be technical or semantic. Data may be stored in different format (e.g. 32 vs. 64 byte, little- vs. big-endian), interface elements with the same semantics may have different names, interface elements with the same name may have differing semantics and so on. One system may return data as a collection, while another one may expect elements of that collection one at a time. Manually detecting and resolving these conflicts is a tedious, error prone work. Often, a lot of glue code is needed to make systems work together [1].

We have developed a prototype for a framework that can help automate some of the steps in conflict resolution. Interfaces and their elements can be semantically described using ontologies in order to facilitate this automation.

In this paper, we describe a prototype for this framework and how we have tested it on a real-world integration scenario.

II. RELATED WORK

In [2] a framework is given for conflict analysis and composition at the component level. Components that originate in object oriented middleware are represented canonically on common denominator basis. The framework is model based. A classification of semantic conflicts is given in [3]. Here, three dimensions are used for classification: naming, abstraction and level of heterogeneity.

One (meta) model-based platform for integration is given in [4] along with the accompanying methodology. It allows for a tight cooperation with the domain expert. The platform enables semi-automatic conflict analysis. An example of using ontologies expressed using the Web Ontology Language (OWL) and available on a network is shown in [5]. It was concluded that by adding classes to an existing ontology, enterprise integration was possible in hours time. An approach called ODSOI (Ontology-Driven Service-Oriented Integration) was proposed for using a combination of ontologies and web services in [6] to address some problems of enterprise integration, along with a vision of an integration framework.

Following Model-Driven Architecture and using a Domain Specific Language (DSL) was proposed in [7]. A DSL called Guaraná was proposed for design and automatic deployment of integration solutions. Another (internal) DSL for enterprise integration called Highway is presented in [8]. It is based on Apache Camel and Clojure. An executable DSL that is platform-independent and message-based is described in [9].

In [10] chapter 10 discusses using Ontology Architectural Patterns in order to achieve semantic enterprise interoperability.

III. SCENARIO DESCRIPTION

To test our approach, we have chosen to recreate an existing integration solution. The scenario is part of a project management portal. The portal is designed to query a web service in order to get data about projects, tasks, etc. However, some legacy data is stored in a SAP system. An integration solution is therefore needed that will expose a web service. Upon getting a Simple Object Access Protocol (SOAP) request for that service it should fetch data from SAP and then wrap the data in a SOAP response. Data is retrieved from the SAP system by placing a Business Application Programming Interface (BAPI) call to a function. A schematic view of the scenario is shown in Fig. 1.

The existing solution was manually coded in Java and uses Hibersap[1] for mapping data from SAP to an object model. It is being executed in the SAP Process Integration

[1] A library for mapping Java classes to SAP backend via annotations, http://hibersap.org

(PI) middleware. Had there been another source of legacy data (e.g. a database) the process of manually writing a mapper would have to be repeated. It could be argued that this scenario is not overly complex and that a simpler solution might have been devised. However, it is important to remember that one of the challenges of developing integration solutions (as for software development in general) are the inevitable changes. For this reason, even for a simple scenario, a flexible solution might present a good investment.

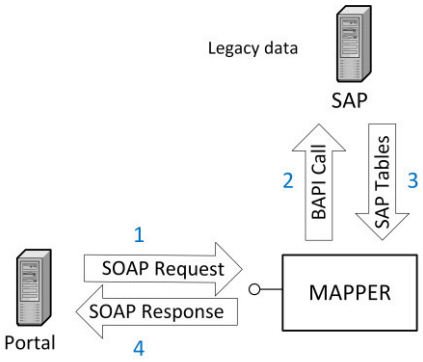


Figure 1. Integration scenario

IV. BASE PLATFORM

Rather than developing a prototype from scratch, we have decided to modify and extend one of the existing IDE tools for integration. We needed a tool that can visually represent the integration solution, be extendable and offer access to the internal object representation of the solution that the user has built, so that we could use it for analysis and code generation. After researching several available platforms, for this purpose we have chosen Talend Open Studio for ESB (TOS). It is a tool that enables users to develop, test, deploy and administrate integration solutions. A user can graphically lay out the integration solution either as a TOS Job (in *Integration view*) or an Apache Camel route (in *Mediation view*). The Mediation view uses standard Enterprise Integration Pattern (EIP) graphical representation established in [1]. Jobs and routes can cooperate. Both jobs and routes can be run either inside TOS or deployed to a standalone runtime environment. Talend's runtime is based on Apache Karaf and Apache Camel. When a job is run from within TOS, real-time performance and error information is available on the graphical editor for each component and connection. The application comes with connectors for a large number (more than 800) of data sources: files, databases, web services, REST, e-mail, open and proprietary protocols, big data, cloud, .NET etc. An SDK is available for component development. TOS has built in code generation (based on Java Emitter Templates), which further shortened our development process. Various orchestration tools are available based on time, system or user events.

Source code for TOS is freely available on GitHub. The architecture of TOS is that of an Eclipse Rich Client Platform. This allowed us to easily extend and modify parts of the application.

V. SOLUTION

The web service definition (WSDL) file was loaded into TOS. This makes the service available and loads service metadata. From the loaded WS definition, a Job can be created automatically that contains TOS components necessary for the WS request and response. The request component can be configured in terms of address and port on which the WS will be available. Accessing data from SAP is done using the tSAPInput component. Connection to the SAP server is configured in the tSAPConnection component. The tSAPInput component makes a Business Application Programming Interface (BAPI) function call to the SAP system and receives several tables as a response, describing the structure of managed projects (name, planed begin and end date, actual duration, related projects, subproject hierarchy etc.). Extraction of the BAPI function metadata (input and output parameters) was done by writing a Python script that converts a TXT file exported from SAP GUI to an XML representation that can be used for loading into TOS. Automated interface extraction for SAP is available as an extension in TOS, but only in the Enterprise version.

The TOS Job is laid out as in Fig. 2. When a request is received for the WS, connection to the SAP server is initiated. If the connection is successful, the SAP input component fetches data by making an SAP BAPI call. Error handling is possible in TOS, but is omitted here for clarity. Returned data is then fed to a tXMLMap component. This component does the necessary mapping between the format in which data is received from SAP and the format in which it needs to be sent as a SOAP/XML response.

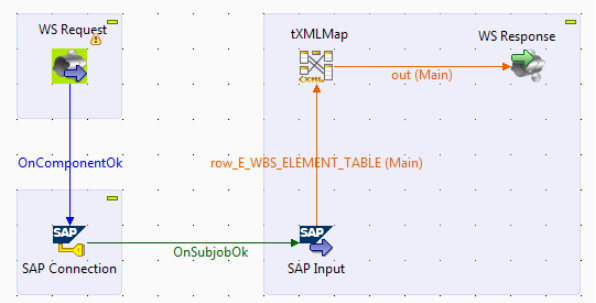


Figure 2. Talend Open Studio job with components laid out and connected

The stock tXMLMap component in TOS has an Auto Map feature, which makes a connection between input and output interface elements if they have the exact same name. Another component, tMap works the same, but accepts different kinds of input and output formats. We have used these two components as a basis for the prototype of our framework.

First, we have extended metadata property editors in TOS. These metadata editors already allow the user to choose name, type, length, date format, etc. for interface elements (rows in TOS terminology). We have added a way for the user to load an ontology file specified in OWL and then to annotate interface elements with one or more ontology elements.

The ontology can be used to formally specify knowledge about systems and their interfaces. This can then be used in the process of conflict detection and resolution. For loading, manipulating and persisting of the ontologies, we are using the Apache Jena framework. In order to use it in the OSGi environment, we've encapsulated Jena as an Eclipse plugin [11].

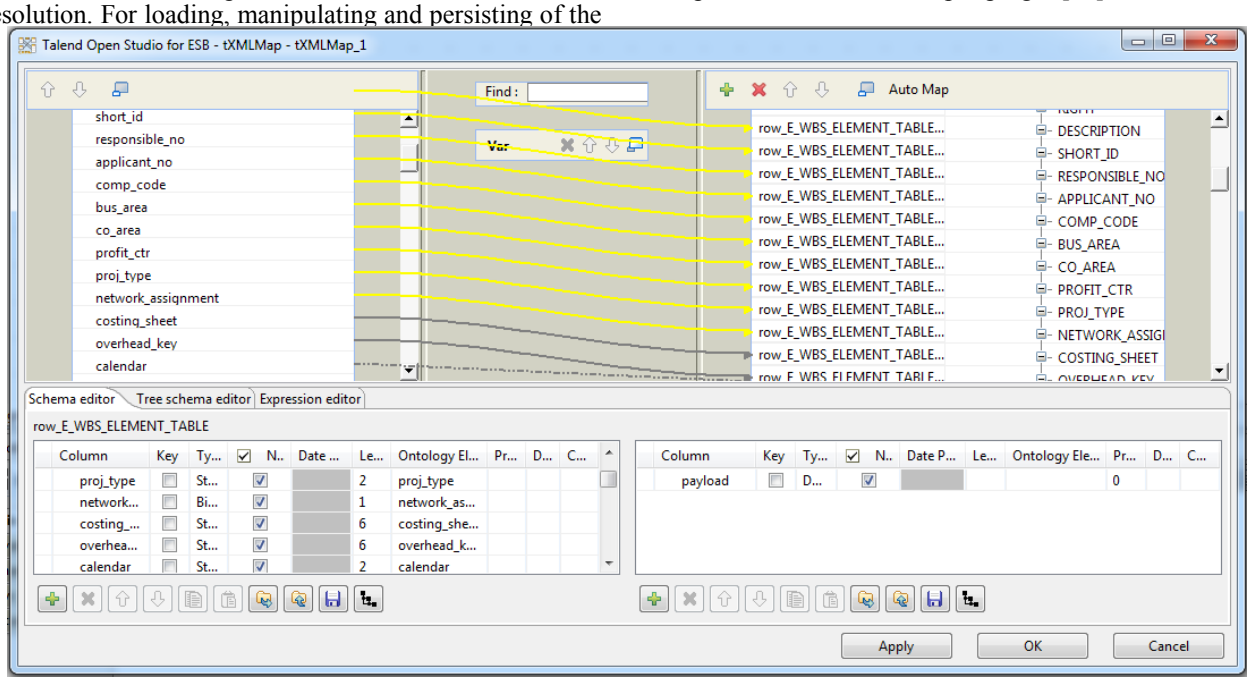


Figure 3. The modified tXMLMap component with automatically mapped input and output interface elements. The button next to the Save button is used to load an ontology.

In general, knowledge about involved systems could be in multiple ontologies and should therefore be merged before they are used. Since ontology merging is not trivial operation and is out of the scope of our research, at this time we assume that it has been already performed and we operate with a single ontology that describes all systems that are involved.

Once interface elements are annotated with ontology elements, this semantic description can be used by the Auto Map feature. Our current implementation is rather simple: it will map those input and output elements that are annotated with the same ontology element. It could be argued that this strategy may offer a marginal time saving in the overall integration process and even take longer, depending on the scenario. It can certainly be said that it does not provide any conflict resolution. However, our current goal was only to establish a solid platform where semantic data will be available in a powerful, real world environment. Further work on developing the actual conflict detection and resolution is to follow.

The mapper is not limited to one input and one output interface. Multiple input and output interfaces can be involved in a single mapping. Likewise, one interface element may be annotated with multiple ontology elements. When an element of the output interface is matched with more than one input element, the way in which those two elements will be merged may also present a semantic conflict. Dealing with these merge conflicts is another point for future research. In the prototype we have opted for the simplest possible solution: data from the input elements is converted to strings and those strings are concatenated.

Figure 4. Elements from multiple input interfaces mapped to a single output interface element

It is important to note that after automatic mapping is performed, the user has the ability to manually review and edit the mappings.

The user interface of the map component can be seen in Fig 3. and Fig. 4. Input interface is on the left and output interface on the right.

VI. FUTURE WORK

During the development of this prototype, a general workflow needed to develop an integration solution became apparent and is shown in Fig. 5. As stated earlier, the first step - processing ontologies is out of the scope of our research. The final step, code generation for target ESB runtimes is already available in TOS. We plan to focus on the three inner steps: finding mappings, detecting and resolving semantic conflicts and building mapping expressions, which are then used by the TOS code generator.

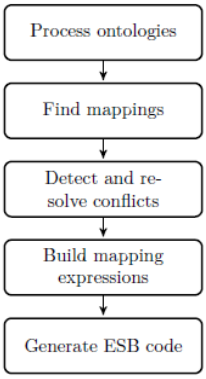


Figure 5. General framework workflow

Given the vast possible kinds of conflicts and the fact that each of them may appear in a different context for each integration scenario, we conclude that developing a fully automated solution for each such case would be next to impossible. For this reason, we plan to develop the framework in such a way that it may be customizable by the user. One way that we see as adequate for this is developing a conflict resolution DSL. Each step of the framework workflow could then be parametrized by the user, without the need for coding in some general purpose language like Java. The user could specify how each type of conflicts will be handled in the given context. A sensible default should exist for each conflict type.

To make the DSL usable and productive, an editor will need to be available for it that offers the usual coding aids to the user: syntax coloring, code completion, instant error checking and highlighting [12]. To develop such an editor and integrate it into our existing prototype, which is Eclipse based, a tool like Xtext or Spoofax/IMP may be used.

VII. CONCLUSION

Extending an existing integration platform has tremendously cut the time needed to develop a working prototype that we can use to test our framework. Using this prototype, we were able to import an ontology that describes the involved interfaces. We have then annotated elements of those interfaces with ontology elements. This information was then used by the Auto Map feature that we have developed to correctly connect elements of the input and output interfaces. After the automatic mapping, the user is able to visually inspect the results, review them and modify if necessary. These mappings were then used to generate code that successfully replaced an integration solution that was previously manually coded and used in a real-world application. Using such a powerful tool, along with the ability to map elements automatically shortens the integration solution development time, while allowing for a very flexible solution that can be modified as necessary if any of the involved systems, or enterprises themselves change with time.

We plan to further research this subject and develop a way for the end user to describe conflict resolution rules for within the context of their own scenario. One way we see fit for this purpose is developing a DSL in which the user could describe how each type of conflict will be handled.

ACKNOWLEDGMENT

Part of the research was done at the offices of PI Informatik, GmbH in Berlin. We wish to extend our gratitude for their help and hospitality.

REFERENCES

[1] G. Hohpe, B. Woolf. “Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions”. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2003.

[2] A. Leicher. “Analysis of Compositional Conflicts in Component-Based Systems”. PhD thesis, Fakulät IV - Elektrotechnik und Informatik der Technischen Universität Berlin, 2005.

[3] A. Ouksel, E. Naiman. “A classification of semantic conflicts in heterogeneous database systems”. Journal of organizational computing, 1995.

[4] R. Kutsche, N. Milanovic, G. Bauhoff, T. Baum, M. Cartsburg, D. Kumpe, J. Widiker. “BIZYCLE: Model-based interoperability platform for software and data integration”. TU Berlin, 2008.

[5] S. Stoutenburg, L. Obrst, D. Nichols, P. Franklin, K. Samuel, M. Prausa. “Ontologies in OWL for Rapid Enterprise Integration“. OWLED 2007 Workshop on OWL: Experiences and Directions

[6] S. Izza, L. Vincent, P. Burlat, “A Unified Framework for Enterprise Integration. An Ontology-Driven Service-Oriented Approach”. Interoperability of Enterprise Software and Applications INTEROP-ESA, 2005

[7] H. Sleiman, A. Sultán, R. Frantz, R. Corchuelo. “Towards Automatic Code Generation for EAI Solutions using DSL Tools”. JISBD, 2009

[8] V. Kovanović, D. Đurić. “Highway: a domain specific language for enterprise application integration”. 5th India software engineering conference. ACM India, 2012

[9] M. Shtelma, M. Cartsburg, N. Milanovic. “Executable Domain Specific Language for Message-Based System Integration”. MODELS 2009, USA

[10] Y. Charalabidis. “Revolutionizing Enterprise Interoperability through Scientific Foundations“. IGI Global, 2014

[11] J. McAffer, J. Lemieux, C. Aniszczyk. “Eclipse Rich Client Platform”. Addison-Wesley Professional; 2 edition, 2010

[12] M. Voelter, “DSL Engineering: Designing, Implementing and Using Domain-specific Languages”, CreateSpace Independent Publishing Platform, 2013

DATA POINT MAPPING APPROACH TO AIRPORT ONTOLOGY MODELLING AND POPULATION

Nikola Tomašević, Marko Batić, Vuk Mijović, Sanja Vraneš
University of Belgrade, Institute Mihajlo Pupin

Abstract – *For development of the intelligent airport energy management system, a comprehensive airport data model is required to describe all static knowledge relevant for the airport energy management, enabling the integration and interoperability of different technical systems. One way of providing such comprehensive airport data model is based on the ontology modelling paradigm. Having in mind that airports are rather complex infrastructures with numerous, heterogeneous devices coming from different vendors, and therefore characterized with the large quantities of the static data, it is important to choose a suitable approach for ontology modelling. This paper presents one of the possible approaches to the airport ontology modelling and population which combines three different, but yet complementary methods: LODRefine tool, SPIN mapping and SPARQL Update queries. The flexibility of the proposed approach was seen in possibility to instantiate any airport infrastructure of the given complexity. The input for population of the airport ontology model was extracted from the various data sources such as data point lists, technical sheets, audits, interviews, questionnaire etc. Finally, as a test-bed platform for the ontology population, two specific airport infrastructures were chosen, Malpensa airport in Milan and Fiumicino airport in Rome.*

1. INTRODUCTION

Present-day airport energy management systems (EMS), are leveraged upon the legacy supervisory control and data acquisition (SCADA) systems and building management systems (BMS) which are faced with the problem of complex heterogeneous infrastructure comprising high number of the field devices produced by different vendors and using different protocols. For providing more advanced airport EMS, classification and description of the target airport facility within a comprehensive airport data model is a prerequisite. This implies description of the belonging devices, systems, their technical characteristics and relations, communication protocols, semantics of the low-level data etc.

One way of providing such comprehensive airport data model is based on the ontology modelling paradigm which was proposed in [1]. Ontology is one of Semantic Web pillars and can be defined as a formal way of representing the knowledge as a set of concepts and their relations in domain of interest. In other words, the ontology is a formal, explicit specification of a shared conceptualization [3]. Therefore, it can be utilized to provide the description of the domain of interest by defining the related entities and their interdependencies, but also to reason upon the modeled entities. The advantage of the ontology concept reflects in reducing the field level heterogeneity and in easier adoption of the future technical equipment. So far, a number of ontology-based facility management frameworks were proposed in the literature such as [4], [5] supporting the integration of multi-vendor devices/sensors. Furthermore, the ontologies were used to increase the energy efficiency and to provide adequate energy saving strategies of so-called smart homes such as in [6], [7]. However, none of them deals with the airport energy management domain and provides a holistic approach to the airport facility modelling.

On the other hand, an ontology-based Airport Data Model (ADAM) layer serving as a common metadata layer of the airport EMS was proposed in [1] and [2]. ADAM layer served as a central data repository which provided the needed semantics to the involved EMS components. In that way, the integration and interoperability of the overall system was supported. The airport ontology developed as part of the ADAM layer, provided a semantic enrichment of low-level signals in order to provide a high-level messages to the airport operator or airport manager. First, a generic model was developed, i.e. the core airport ontology, which had to be extended and instantiated further to model a specific airport infrastructure.

This paper provides a detailed explanation of the ontology modelling procedure and presents one of the possible approaches which could be undertaken for ontology population task. At the start of the development procedure, it was necessary to identify a suitable modelling approach and to choose general concepts behind the modelling. This influenced the decision about the ontology model characteristics such as granularity, abstraction and classification of related entities within the ontology class hierarchy.

As previously mentioned, the airport ontology served as a central data repository of the overall airport EMS solution, and therefore it had to be populated with the static data about the airport facility and target systems/equipment (such as significant energy consumers). First, it was needed to acquire the relevant semantics and to populate it within the airport ontology model. This implied the definition of new concepts, as well as instances and their properties. As proposed in [1], the core airport ontology was first defined providing a generic model of the airport facility. To model a specific target airport infrastructure, the core airport ontology had to be extended and populated based on the static data

gathered from the field. For both modelling and populating the airport ontology, two specific airport infrastructures were used as a test-bed platform, Malpensa (MXP) airport in Milan and Fiumicino (FCO) airport in Rome due to their rather complex infrastructure and variety of the field-level devices installed at the site.

Input for the ontology population was extracted from the raw data points (i.e. low-level signals) monitored by the BMS which were identified as relevant from the energy management perspective. This data were used first to extend the core airport ontology model and then to populate it. In this way, it was possible to instantiate any airport infrastructure of interest which is the main advantage of the proposed approach. This paper also elaborates in detail possible data point mapping approaches which were used for the airport ontology population. This task was mainly performed in fully automated manner. Three different, but yet complementary approaches have been taken into account for translation of the data point lists into the ontology: LODRefine tool [8], SPIN mapping [9] and SPARQL Update queries [10]. To provide an overview of their functionalities and capabilities, each of the mentioned approaches were described in detail.

The reminder of this paper starts with Section 2 in which the main aspects and objectives of the airport ontology in context of airport EMS are described. Section 3 elaborates in detail the overall airport ontology modelling procedure and some modeling issues which influenced the structure of the airport ontology hierarchy. Section 4 presents possible data point mapping approaches which were utilized to perform the population of the ontology model. Finally, Section 5 concludes the paper.

2. AIRPORT ONTOLOGY SPECIFIC ASPECTS

Before detailed elaboration regarding the airport ontology development and population approach, a brief overview of the general context in which the airport ontology is used as a meta-model and platform supporting the integration and interoperability between different subsystems is given. As previously mentioned, a need for a common metadata layer resulted in the development of a unique Airport Data Model (ADAM) (shown in Figure 1) for which an ontology-based approach was adopted [1], [2]. The aim of the ADAM layer was to serve as a common knowledge base repository used by the EMS components and to contribute to the integration and interoperability of these components. It provides a semantic enrichment of various signals coming either from the legacy BMS, SCADA, applied fault detection and diagnosis (FDD) algorithms or directly from the sensors/data-loggers installed within the airport facilities, thus enabling the high-level information for the end-user (airport operator or energy manager) as described in [1].

The main objectives for the development of ADAM, tailored to suit the needs of the comprehensive EMS solution, would be the following:

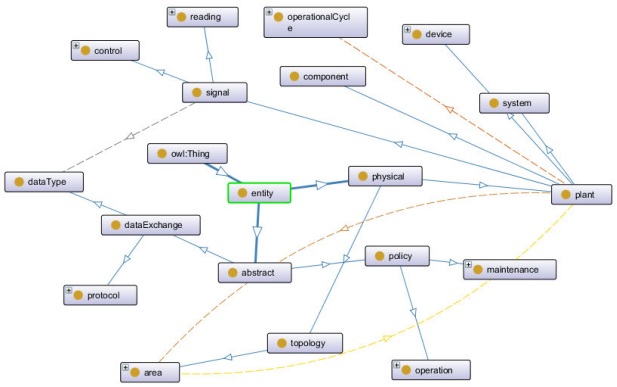


Figure 1. Excerpt of airport ontology class relations and dependencies.

- **Modelling the domain of interest**, i.e. defining the infrastructure of the concrete airport building by classifying installed systems relevant to the energy consumption aspect with belonging sensor/actuator devices;
- Providing means for **technical characterization and semantic interpretation** of signals going to/from the installed equipment (which incoming/outgoing signal belongs to which device/system, what its characteristics are, relations to other devices/signals);
- Providing the **topological profile of the airport facility** and information about geographical location of every installed device/signal (useful for analyzing spatial correlation of data).

Initially, it was important to identify the modelling approach which will be undertaken and the general concepts behind the modelling. The decision regarding the granularity, abstraction and classification of different entities at different levels of the ontology hierarchy is highly influenced by these issues. Since it was meant to be used as a central data repository of the airport EMS, the airport ontology had to be populated in order to store all the static data regarding the targeted systems (information related to the Significant Energy Users (SEU), e.g. air handling units (AHU), such as nominal mass flow, fan drive power etc.) which will be later used to calculate the energy waste due to faults and corresponding saving potential. In other words, for development of the ontology-based ADAM layer, first it was necessary to acquire all relevant information about the airport infrastructure and then to transfer it into the airport ontology which affected both the domain model (ontology hierarchy) and creation of new instances of the ontology entities. For transferring the gathered information into the airport ontology, i.e. for the task of the ontology population, it was important to select yet a suitable approach which will be elaborated in detail in Section 4.

3. AIRPORT ONTOLOGY MODELLING

At the beginning of the airport ontology development process (as shown in Figure 2), the core airport ontology was defined providing a generic model of the airport

facility [1]. Core airport ontology was comprised of the common concepts identified as relevant from the perspective of the energy management usually present in the airport infrastructures. It provided modelling guidelines to describe the technical characteristics and topological profile (precise location) of the systems and devices installed at the site.

As a test-bed platform for the ontology modelling, two specific airport infrastructures were chosen, Malpensa (MXP) airport in Milan and Fiumicino (FCO) airport in Rome. Serving as a two major European air-traffic hubs, MXP and FCO airports were taken as suitable also owing to their rather complex infrastructure, technical characteristics and different aspects of the existing devices and modules. Based on the data gathered for the technical characterization of the ICT systems and the analysis of the different aspects of the involved devices and equipment, the core airport ontology had to be extended and further instantiated to model specific target airport facility. In other words, by extending and populating the core airport ontology these two separate ontology instances emerged representing two full-blown airport ontology models tailored according to the current state of the target pilots.

Development of the airport ontology (Figure 2) consisted of the definition of the main ontology entities with the corresponding properties and relations, as well as the harmonization with the contemporary ontology modelling regulations and standards. For this purpose, the leading ontology modelling standards (SUMO ontology [11], IFC (Industry Foundation Classes) data model [12] and CIM as part of IEC 61970 series of standards [13]) were taken into account. For ontology modeling, OWL as one of the most utilized modeling languages was used [14]. Protégé™ tool [15] was selected as ontology development framework used not only for modelling but also for ontology population tasks.

An excerpt from Protégé user interface depicting the water pump entity as part of the extended airport ontology model is presented in Figure 3. It could be seen from Figure 3 that each instance of the class "waterPump" is modelled with belonging list of properties such as "device_id" representing the unique identifier (ID) of the device, "partOf_system" indicating the system to which this specific device belongs, "belonging_signal" representing the list of belonging "signal" instances etc. Furthermore, some of the nominal technical characteristics of device are represented by properties such as "outputPower_kW", "hydraulicPressure_kPa" and "waterFlow_l/s". The topological perspective of the corresponding device instance is modelled with the properties representing the area/position where that specific device is installed (properties "locatedAt_area" and "position").

The input for population of the airport ontology model was extracted from the various data sources such as BMS

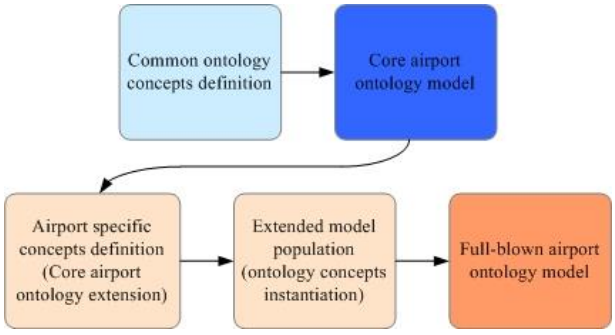


Figure 2. Airport ontology development process.

data point lists, technical sheets, equipment manuals audits, interviews, questionnaire etc. These sources contained the information that enabled classification of energy management related entities (such as HVAC system, power supply, water supply system etc.) and definition of the corresponding relations between them. The extracted information had to be transferred into the airport ontology by instantiating the corresponding instances of the ontology entities with belonging property values and relations. Once the airport ontology was populated, semantic queries were used for extracting the knowledge from the airport ontology.

Both extension and population of the core airport ontology had to be performed based on the semantics of the raw data coming from the field devices, i.e. low-level signals. The data semantics were extracted from the BMS data point lists carrying the information about every low-level signal that the EMS might encounter. Since at both target airports, the semantics of the low-level data/signals were aligned with the Unified data point naming (UDPN) convention [16], harmonization of the airport ontology instances was carried out correspondingly. The harmonization considered mapping of the data point semantics, such as signal identifier, signal source, data type, signal characteristics etc. into the airport ontology as corresponding entity instance or property value. Thus, the flexibility of such ontology modelling and population approach could be seen in possibility to instantiate any airport infrastructure starting from the core model. The following section will elaborate in detail possible data point mapping approaches which were used to populate the airport ontology.

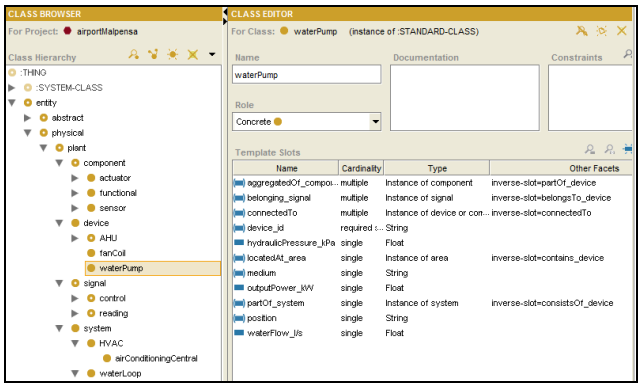


Figure 3. Device entity (water pump) properties.

4. DATA POINT MAPPING APPROACH

For providing a way for common understanding among EMS components, as already mentioned, population of the airport ontology had to be performed in correspondence with the UDPN convention [16] utilized for raw data/message exchange. More precisely, ontology population was carried out based on the BMS data point lists holding raw data semantics compiled according to the UDPN convention. In order to transfer the raw data semantics into the airport ontology, each entity instance was defined with the property values representing the corresponding UDPN attributes. For example, class "signal" instance was defined with properties representing the location of the source device (UDPN attributes "building" and "zone"), (sub)systems to which it belongs (UDPN attributes "system" and "subsystem1/2"), signal type (UDPN attributes "medium", "position" and "kind") and unified data point name representing unique signal ID. In that way, both extension and instantiation of the core airport ontology into two full-blown airport ontology instances (for MXP and FCO airports) was carried out in line with UDPN convention.

In order to extend and populate the airport ontology model according to the UDPN convention, BMS data point lists (i.e. Excel sheets carrying the information about every low-level signal/device that EMS might have to deal with), shown in Figure 4, were mapped into the airport ontology entities. More precisely, the data point lists were used first to extend the core airport ontology model by developing more fine-grained model. This implied identification of sub-concepts of already existing ones, such as definition of different types of sensors/actuators, signals etc. After the core airport ontology was extended, population of model had to be performed. This considered mapping of the raw data semantics, i.e. the corresponding data cells (representing previously mentioned UDPN attributes) from the data point lists into the airport ontology as corresponding entity instance or as a property value of specific instance (such as properties of class "signal" instance).

Population of the information into the airport ontology was mainly performed in fully automated manner, while the "fine tuning" (for instance, establishment of additional relations among newly created and/or mapped instances/properties) was performed subsequently in a semi-automatic manner. These alignment tasks had to be carried out for both airport ontology instances representing the full-blown facility models of MXP and FCO airport.

Having the previously mentioned in mind, the following three different, but yet complementary approaches have been considered for mapping of the information into the airport ontology:

- LODRefine tool [8],
- SPIN mapping [9], and
- SPARQL Update queries [10].

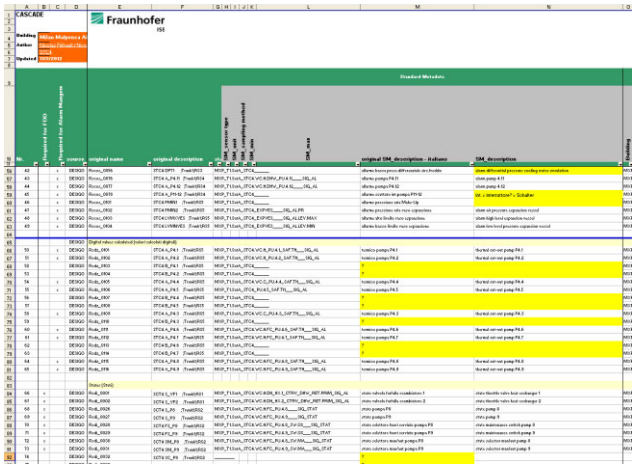


Figure 4. Excerpt from data point lists.

Having in mind their individual functionalities and advantages, these approaches could be implemented separately or in combination. However, for population of the airport ontology to model both MXP and FCO airport facilities, they were applied in combination in order to exploit their full capabilities. To provide the qualitative overview of their particularities, each approach was elaborated in the following subsections.

A. LODRefine transformation

LODRefine [8] is a LOD-enabled version of OpenRefine tool for cleaning, linking and transformation of data from one format into another. It is a part of the LOD2 Stack, which is the output of the FP7 project LOD2 (Grant Agreement No. 257943) and it is comprised of tools for managing the life-cycle of Linked Data. For the purpose of this paper, LODRefine tool was utilized for automatic translation of information stored within BMS data point lists (in form of Excel sheets for both MXP and FCO airports) directly into the RDF triplets simply by defining the translation rules (using the corresponding LODRefine user interface functionalities) as it can be seen in Figure 5.

For definition of data extraction/translation rules, i.e. for establishment of the target RDF triplets, entities and corresponding properties/relations of the core airport ontology were utilized as a base. The result of data translation/mapping is shown in Figure 6 in a form of RDF triplets carrying the extracted semantics (such as signal ID represented in compliance with the UDPN convention, indicated by a red rectangle). Finally, the RDF triplets carrying the extracted information regarding the newly created instances and their belonging property values were subsequently imported (as OWL ontology model carrying the extracted semantics) into the core airport ontology model using the Protégé editor.

In this way, the LODRefine tool was first utilized to perform extension of the core airport ontology and then for the population of the extended ontology model. At the same time, this approach provided the possibility to perform the "raw" mappings and alignment with the

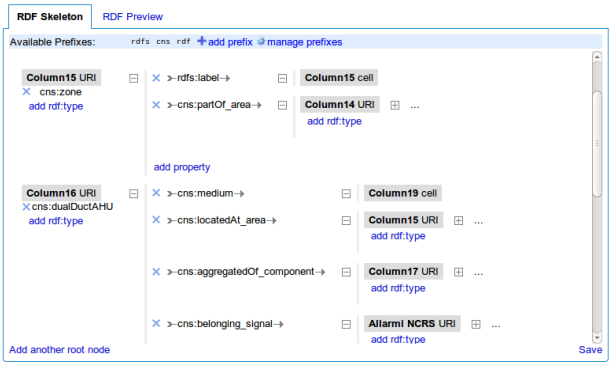


Figure 5. LODRefine translation rules.

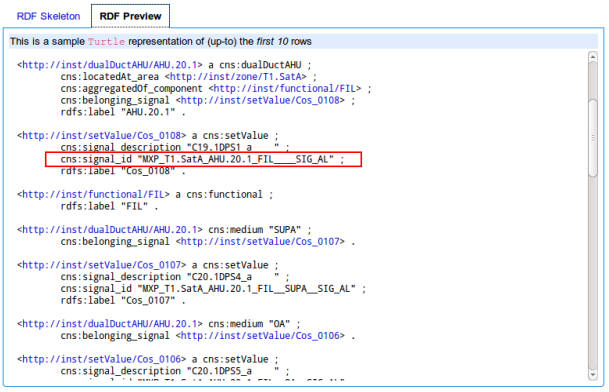


Figure 6. Extracted semantics as RDF triplets (LODRefine)

UDPN convention in completely automatic manner, while the following two approaches were considered subsequently as more convenient for "fine tuning" of the airport ontology in semi-automatic manner as it will be described further.

B. SPIN based mapping

SPARQL Inferencing Notation (SPIN) [9] is a SPARQL-based language used for representation of the mappings between RDF/OWL ontologies. It is used to transform instances of source classes into instances of target classes. Furthermore, it is used for definition of rules that map not only the entire instances but also their specific properties/values. For definition of mapping rules, SPARQL CONSTRUCT keyword was used to map classes from one graph pattern to another while binding the source and target classes was performed in the WHERE clause of the SPIN mapping rule shown below.

```
CONSTRUCT {
    ?target ?targetPredicate1 ?newValue .
}
WHERE {
    ?this ?sourcePredicate1 ?oldValue .
    BIND (spin:eval(?expression, sp:arg1, ?oldValue) AS ?newValue) .
    BIND (spinmap:targetResource(?this, ?context) AS ?target) .
}
```

Following this approach, instances and corresponding relations/properties had to be initially extracted (such as from the BMS data point lists) and imported as RDF

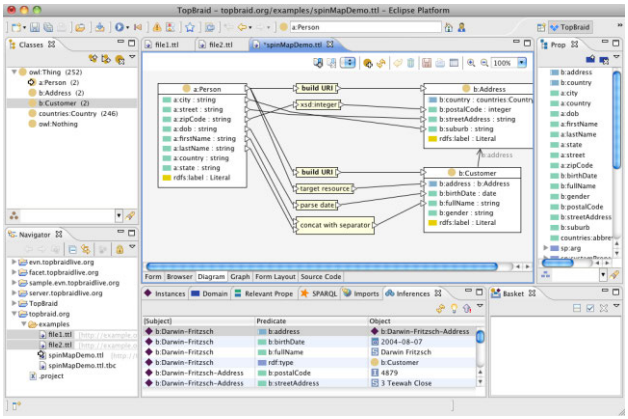


Figure 7. SPIN based mapping.

triplets (simply by importing the corresponding CSV file carrying the source information) that could be further mapped (taken as a source objects with corresponding property values) into the target model, i.e. the entities of the previously extended airport ontology model (by using the SPIN mapping rules). Based on the corresponding pre-defined mapping rules, desired relations were established, not only among already existing instances (generated by using the first approach), but among the newly imported ones as well.

Additionally, SPIN mapping language is a high-level language that is suitable to be edited graphically (simply by drag and drop of source/target classes). This convenient feature was provided by TopBraid Composer starting from release 3.5. Furthermore, it provides a visual editor (shown in Figure 7), i.e. graphical user interface, that significantly facilitated alignment tasks of the airport ontology entities and their interdependencies.

C. SPARQL Update based mapping

Final "fine tuning" of the airport ontology model was performed through the SPARQL Update [10] queries performed from the Java script directly upon the ontology model. SPARQL Update is an extension of the SPARQL query language that provides the ability to add, update and delete RDF data. It enables the generation of specific SPARQL Update queries including commands such as MODIFY (for modifying the existing instances), INSERT (for inserting the newly created instances/property values) and DELETE (for deleting the existing instances/property values) based on which target entity property can be updated according to the handled data. New instances and property values which should be inserted/modified in the ontology model could be manually defined either within the Java script itself (through corresponding query arguments) or extracted from the source Excel sheet handled by JAVA Excel API.

This approach allowed "semi-automatic" handling of the already created ontology instances, i.e. it provided the possibility for an update of the target instances and corresponding properties/relations of airport ontology. For performing alignment and mapping tasks, patterns of SPARQL update queries were defined based on which

corresponding modifications were carried out (within the JAVA script). Example of SPARQL Update query is shown below. From the presented example, it could be seen that property "waterFlow_l/s" of the target device "waterPump", having the specific device ID (taken as the input argument "id") , was updated (more precisely, first deleted and then updated) with the new nominal water flow data (taken as an input argument "flow").

```
// SPARQL Update - To replace a property first delete current value and then
insert new one
queryStr.append("DELETE { " +
            "?s pref:waterFlow_l_s ?o. "+
            "} " +
            "WHERE { " +
            "?s pref:device_id \"" + id + "\"^^xsd:string. " +
            "?s pref:waterFlow_l_s ?o. " +
            "} " +
            "INSERT { " +
            "?s pref:waterFlow_l_s \"" + flow + "\"^^xsd:float. " +
            "} " +
            "WHERE { " +
            "?s rdf:type pref:waterPump. " +
            "?s pref:device_id \"" + id + "\"^^xsd:string. " +
            "}");

// Update execution
UpdateAction.parseExecute(queryStr.toString(), model);
```

5. CONCLUSIONS

To provide advanced and more intelligent airport EMS, an ontology-based ADAM layer was proposed in [1] and [2]. The airport ontology, developed as part of ADAM layer, served as a central data repository which provided the needed semantics to the EMS components, thus supporting the integration and interoperability of the overall system. First, the core airport ontology providing a generic model of the airport facility was modelled, which had to be extended and instantiated further to model a specific airport infrastructure. This paper proposes one of the possible approaches to perform the airport ontology modelling and population. In other words, it explains the undertaken modelling approach and the general concepts behind the modelling.

The task was to populate the airport ontology with the static data regarding the airport facility and target systems/equipment (such as significant energy consumers). To provide the input for modelling, first the needed data and relevant semantics were acquired. Then, the core airport ontology had to be extended and populated based on the acquired data to model a specific target airport infrastructure. This included the definition of new concepts, instances and their properties. Two European airports were taken as a test-bed platform for modelling tasks, MXP airport in Milan and FCO airport in Rome. For population of the airport ontology, the BMS data point lists carrying the semantics about the low-level signals were taken into account. Three different, but complementary approaches for translation of the data point lists into the ontology were elaborated: LODRefine tool, SPIN mapping and SPARQL Update queries. An overview of the functionalities for each of the mentioned approaches was provided. In this way, it was possible to instantiate any airport infrastructure of interest.

ACKNOWLEDGEMENTS

The research presented in this paper is partly financed by the European Union (FP7 CASCADE project, Pr. No: 284920), and partly by the Ministry of Science and Technological Development of Republic of Serbia (SOFIA project, Pr. No: TR-32010).

REFERENCES

[1] Tomasevic N., Batic M., Vranes S., "*Ontology-enabled airport energy management*," ICIST 2013, 3rd International conference on information society technology, pp. 112-117, Kopaonik, 2013.

[2] Batic M., Tomasevic N., Vranes S., "*Ontology API for web-enabled FDD system*," ICIST 2013, 3rd International conference on information society technology, pp. 142-147, Kopaonik, 2013.

[3] Gruber T.R., "*A translation approach to portable ontology specifications*", Knowledge Acquisition, Vol. 5 (2): pp. 199–220, 1993.

[4] Dibley M., Haijiang Li, Miles, J., Rezgui, Y., "*Towards a synchronized semantic model to support aspects of building management*", 7th IEEE International Conference on Industrial Informatics, Cardiff, Wales, pp. 307-312, 2009.

[5] Praus F., Granzer W., Kastner W., "*Enhanced control application development in Building Automation*", 7th IEEE International Conference on Industrial Informatics, Wales, pp. 390-395, 2009.

[6] Reinisch C., Kofler M.J., Iglesias F., Kastner W., "*ThinkHome Energy Efficiency in Future Smart Homes*", EURASIP Journal on Embedded Systems, 2011.

[7] Rossello-Busquet A., Brewka L.J., Soler J., Dittmann L., "*OWL Ontologies and SWRL Rules Applied to Energy Management*", 13th International Conference on Computer Modelling and Simulation, Cambridge, pp. 446-450, 2011.

[8] LODRefine tool, Available: http://code.zemanta.com/sparkica/

[9] Knublauch H., Hendler H.A., Idehen K. (eds.), "SPIN - Overview and Motivation", W3C Member Submission 22 February 2011, Available: http://www.w3.org/Submission/2011/SUBM-spin-overview-20110222/

[10] Gearon P., Passant A., Polleres A. (eds.), "SPARQL 1.1 Update", W3C Recommendation 21 March 2013, http://www.w3.org/TR/sparql11-update/

[11] Suggested Upper Merged Ontology (SUMO), http://www.ontologyportal.org/

[12] Industry Foundation Classes (IFC) data model, buildingSMART, http://www.buildingsmart.org/standards/ifc/

[13] Common Information Model (CIM) Users Group, http://cimug.ucaiug.org/default.aspx/

[14] Lacy L.W., Owl: Representing Information Using the Web Ontology Language, Trafford Publishing, 2005.

[15] The Protégé Ontology Editor and Knowledge Acquisition System, http://protege.stanford.edu/

[16] Keane M., Costa A., Blanes L., Donnely C., Torrens I., Monaghan P., Brogan M., McCaffrey M., CASCADE ICT for Energy Efficient Airports, Project Deliverable D2.1 – "*CASCADE Methodology for Energy Efficient Airports*", 2012.

Enabling Customization of Document-Centric Systems Using Document Management Ontology

R. Molnar*, S. Gostojić*, G. Sladić*, G. Savić*,Z. Konjović*
*University of Novi Sad/Faculty of Technical Sciences, Novi Sad, Serbia
{rmolnar, gostojic, sladicg, savicg, ftn_zora}@uns.ac.rs

Abstract - This paper introduces a conceptualization of the document management domain that is serving as a foundation for a semantically-driven document management system. The conceptualization is based on the ISO 82045 family of standards for document management and is specified in OWL. Legislative documents were used as a case study and a proof of concept of the proposed conceptualization.

I. INTRODUCTION

As the usage of Document management systems (DMS) [1,2] increases throughout different sectors of the economy, the need for a cost-effective yet domain-specific DMS arises. DMS should enable simple capture, storage, transfer, retrieval and browsing of documents and provide services such as metadata capture, security, integration and version control [3,4]. However, most of the current DMS implementations do not offer domain-specific services (they lack domain semantics) since they are difficult to customize to a particular domain.

Introduction of semantic technologies [5] into document management systems can mitigate those shortcomings by providing an abstract domain-independent model which can be easily adapted to a concrete domain and serve as a foundation for the semantically driven document and workflow management system described in [6].

The semantic model (i.e. ontology) identifies common features of documents belonging to different domains. The fact that domains and documents belonging to those domains differ in their characteristics does not affect the complexity of the proposed model because only the common features, such as document types, document structure, metadata, classes, and identifiers, were modeled. Most of the concepts are co-opted from the ISO 82045 family of standards [2] which is de facto and de jure standard in the document management domain. The model is specified in OWL DL dialect [7] in order to provide the maximum expressiveness possible while retaining computational completeness. The flexibility of the abstract model allows extension with concrete domain-specific elements. Those elements enable implementation of domain-specific services offering additional features. At the same time, the existing features are not compromised, and there are no new constraints on generic documents. A piece of legislation (i.e. a law) was used as a case study of the flexibility of the proposed model. Some important domain-specific services in the legislation domain, which are based on the concrete domain-specific elements, are judgement anonymization and retrieval and browsing of legislation.

II. RELATED WORK

MAchine Readable Cataloging (MARC) [8] is a bibliographic standard that applies to document management by representing the document metadata in a machine-readable form. It is primarily designed to enable exchange of documents between the systems. This standard is an outdated version of the archaic card catalogs. Today's information needs are more demanding than before as there is more than one type of media to be described and managed.

Digital Object Identifier (DOI) [9] framework is used to identify digital objects. It features persistence, network accessibility and interoperability with other systems. It is used by other systems that provide domain-specific identifiers, such as CrossRef [10] and DataCite [11]. The DOI system was initiated in 1998 and has later been standardized as ISO 26324. The DOI can be used to identify any digital object, but the major disadvantage of this system is difficult registration of a DOI name. Unique identification of a digital object is done by Registration Agency and it is not free.

Metadata Encoding and Transmission Standard (METS) [12] is a specialization of MARC standard. METS uses Uniform Resource Identifiers (URI) to identify components that, along with relationships between them, can form one digital entity. Unfortunately, the flexibility of the standards can cause many problems with interoperability as the very same digital object can be represented in many different ways. Such approach makes difficulties in basic information retrieval operations like indexing and searching.

MARC, DOI and METS mostly deal with management of metadata.

MoReq2010 [13] is a comprehensive specification of functions, services and processes that records management systems should support. The specification does not specify which algorithms should be used, but it requires compliance to various registry formats. Records management systems focus on ensuring authenticity, integrity, usability, and reliability to make sure that the records are always available and kept as long as it is necessary. This specification can be used as a very good guideline for the creation of a well-formed electronic document management system. Since mid-2011, when the specification has been published, only one software product has passed a compliance testing and is MoReq-certified.

Ontalk [14] is a document management and retrieval tool that includes semi-automatic metadata generator and an ontology-based search engine. It is based on three on-

tologies: the document schema, the document type, and user domain ontology. The system does not cover document classification and is platform-dependent due to the technologies used to implement it. Also, the search engine works only with properly annotated documents that have to be annotated by the user.

ISO 82045 family of standards defines document management concepts and establishes document management principles covering all the phases starting from the conceptual idea of the document to its deletion. Orthogonal features, such as version control and security, are also in the scope of the standard. Part 1 of the standard is intended to be supported by computer-based systems such as DMS or Product Data Management Systems (PDMS).

III. Document Management Ontology

The reviewed document models are not flexible enough to allow multiple extensions with domain-specific rules. Furthermore, most of them do not introduce the concept of versioning and lack the support for document life-cycle management without which it is not possible to achieve management of documents.

The proposed ontology is strongly influenced by the ISO 82045 family of standards and imports time and provenance related concepts from Time Ontology [15] and PROV-O [16] respectively. It is designed to support standard DMS function and enable easy integration with business processes implementing document life-cycle.

ISO 82045 family of standards defines concepts such as the document, the part of a document, document metadata, the document version, the document relationship, the identification of the document, and the classification of a document. Those concepts and their relationships are shown in Figure 1. It is possible to choose desired level of detail by using only necessary concepts (i.e. neglecting the concepts that are not of interest to the user at the given time).

Each document is an instance of exactly one document type that can be inferred from the relationships it has with metadata and other documents. Document types, defined in plain text and OWL as fragments of expressions written in Manchester syntax [17], are as follows:

SingleDocument (Listing 1) –An identified object associated with metadata whose content is entirely contained within the object (e.g. a note, a picture).

```
Document and
hasFragment some DocumentFragment and
hasMetadata some Metadata and
hasPart exactly 0 Thing
```

Listing 1. Single document

CompoundDocument (Listing 2) –An identified object associated with metadata that has content and contains another document without associated metadata (e.g. a report showing a table or a graph). As defined in [2], a compound document is a document containing documents (parts) that cannot be separately identified and cannot be independently managed.

```
Document and
hasFragment some DocumentFragment and
hasMetadata some Metadata and
isComposedOf (Document and hasMetadata
exactly 0 Metadata)
```

Listing 2. Compound document

AggregatedDocument (Listing 3) – An identified object associated with metadata that may have content and contains other documents with associated metadata (e.g. the web site). In other words, an aggregated document is a document containing separately identified documents (parts) that are logically dependent but can be physically independently managed.

```
Document and
hasFragment onlyDocumentFragment and
hasMetadata some Metadata and
isAggregatedOf (Document and hasMetadata
some Metadata)
```

Listing 3. Aggregated document

DocumentSet (Listing 4) – An identified object with associated metadata that has no content (all content is contained in other documents that are part of the document set). The relationship between the documents is implemented in the same manner as in the *AggregatedDocument*. As defined in [2], a document set is a collection of documents that are managed together as a unit for a specific purpose.

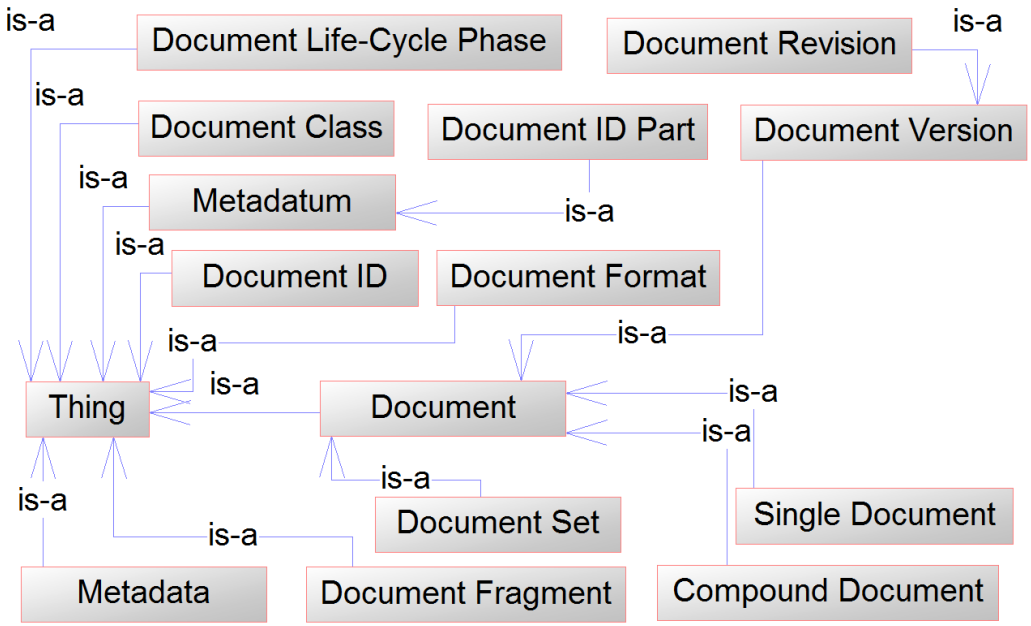


Figure 1.Document management concepts and their relationships

```
Document and
hasFragment exactly 0 Thing and
hasMetadata some Metadata and
isAggregatedOf (Document and hasMetadata
some Metadata)
```

Listing 4. Document set

Properties isAggregatedOf and isComposedOf are subproperties of the hasPart property. Composition relates compound document to their parts and is used when parts are not identified and managed independently. Aggregation relates aggregated documents to their parts and is used to aggregate individual documents into a group which can be described with an additional content and metadata (similarly to a document set that cannot have its own content).

DMS should also support metadata management. Although documents do not necessarily have associated metadata, the proposed ontology supports attaching metadata belonging to various schemata to documents. Since metadata management is out of the scope of this paper, the part of the ontology that deals with metadata is simplified as much as possible. *Metadata* class is implemented as a collection of individual metadata entries. Those entries are individuals of *Metadatum* class that represents a key-value pair.

Each document should be uniquely identified, regardless of whether it is created by the system or another system in its environment. Identification mechanism can be simple or complex, depending on the needs of the user or external identification mechanisms. Each identifier is composed of one or more parts that are individuals of *DocumentIDPart* class, a subclass of *Metadatum* class. Dublin Core Metadata Initiative (DCMI) [18] is just one of many standards proposing that identifiers should be viewed as metadata.

Document classification is another essential feature of document management. Since documents may belong to one or more classes simultaneously (*DocumentClass* class), there is a need to rely on external classification systems. One of the many possible systems is described in [19].

Document content(or document version content) can be either unstructured or structured. In the first case, the content is contained within the document itself. In the other case, the content is contained within document fragments (the individuals of *DocumentFragment* class). The document is structured by defining a hierarchy (isFragmentOf or hasFragment properties) and order (isAfter or isBefore properties) among the fragments. Although those properties should be transitive, due to the computational complexity of the resulting ontology (it would be out of the scope of OWL DL dialect) they are not implemented as such.

The*DocumetnFormat* class is used to specify document format. A document or a document version can be represented in multiple formats (e.g. Microsoft Office Word, PDF, and HTML).

The Time Ontology is an ontology of temporal concepts. The ontology was created as a mean of unifying time-related data that can be found on the internet. The main class of this ontology is the *TemporalEntity* that has two subclasses: the *Instant* and the *Interval*. The *Instant* represents an exact moment in time while the *Interval* represents a time interval. Each interval is determined with its beginning and its end (it can be described as two individuals of the class *Instant*). The interval can also be positively or negatively infinite. A negatively infinite interval is an interval without the beginning, and a positively infinite interval is an interval without an end.

The PROV-O is another World Wide Web Consortium (W3C) recommendation that enhances functionality and interoperability of systems by introducing classes, properties and restrictions used to represent and exchange provenance information. Provenance is defined as information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness [20]. Three main classes of the model are *Agent*, *Activity,* and *Entity*. Their relationships are shown in Figure 2.

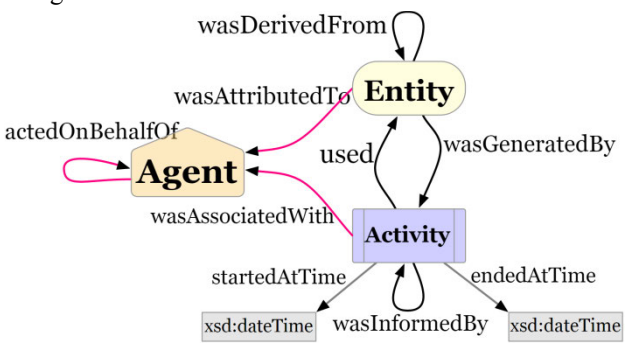


Figure 2. Main PROV-O concepts [16]

The Time Ontology and the PROV-O ontology are used because they are W3C recommendations and have been proven in practice.

Depending on the definition, a new version of the document does not necessarily imply that its content has been changed. It might imply that its presentation did. Since the proposed ontology does not cover the document presentation layer, we opted for an approach in which two document versions differ only by the data they contain. In order to enable versioning of documents, the *DocumentVersion* class is defined as the subclass of the *Entity* class (an entity is a physical, digital, conceptual, or another kind of thing with some fixed aspects). Since document versions also pass trough most of the life-cycle phases as documents themselves, the *DocumentVersion* class is also a subclass of the *Document* class.

In the case of sequentially effective versions, the latest released document version is the only operative, and it serves all intended purposes of all previous document versions. Nevertheless, a document may have more than one effective version at a time (Figure 3). Concurrently effective versions assume that multiple versions of the document are operative, at a particular moment in time, and each effective version still serves its defined purpose [2].

DocumentRevision, which is a subclass of *DocumentVersion* class, is defined as an officially confirmed document version.

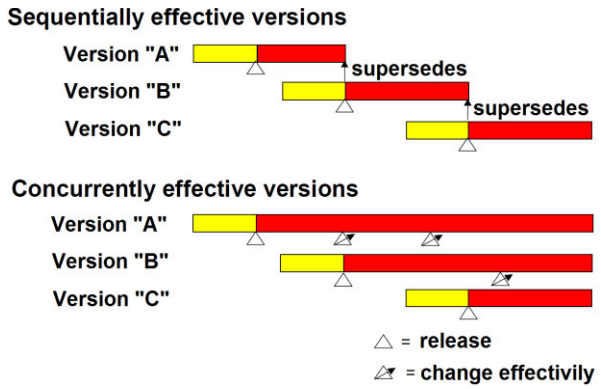


Figure 3. Version effectivity [2]

Document life-cycle has multiple phases. Each of the phases represents a state of the document in time. There are seven phases: Initiation, Preparation, Establishment, Use, Revision, Withdrawal, and Deletion [2] (Figure 4).

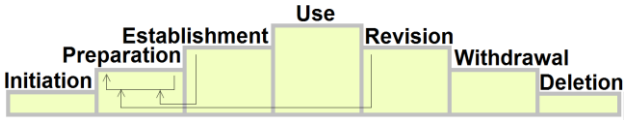


Figure 4.Document life-cycle phases [2]

In the initiation phase, the document is uniquely identified and classified within the system. DMS may use its subsystems or external systems in order to properly identify and classify the document.

The preparation phase is reserved for document content development after which it enters an establishment phase where the document undergoes various checks and approval within the responsible organization. When a document enters the approval process, all changes are traceable, and the document should already be under version control.

The use phase starts as soon as the document is completely verified and released.

During the revision phase, the document content is being changed.

After the document becomes useless, it is then withdrawn. The document itself will be kept for a minimum legally required period (that may vary) as an archive.

Deletion phase is the last phase in document life-cycle, and it means that the document is being completely deleted, and it can no longer be traced. In some specific cases, complete elimination is not possible as there are active references to the document. In such cases, the document metadata is kept only in order to keep the references correct [2].

IV. Case Study

Legal profession is a sector of the economy that uses huge quantity of documents. Apart from having complex structure and being interdependent, those documents are characterized by strict identification mechanisms, metadata and classification schemata, and changing life-cycle management processes. Therefore, we decided to use legislation (statutes, laws) as a case study and proof of concept of the extension and instantiation of the proposed ontology. The abstract document management ontology and its instantiation in the legislative domain are available at [21].

As an example, we instantiated "Law on Personal Data Protection" [22] in two versions (as enacted in 2008 and 2009) in several different formats. Since the legislation is highly structured, the content of those versions is associated with *DocumentFragment* instances that are ordered in the proper manner.

Each law passes through several life-cycle phases before it is used.

In the initiation phase, the bill gets a unique identifier and one or more classifications within some classification system.

There are several mechanisms to identify the bill. The mechanism used in Serbian legislature relies on its name and the volume and the number of the official journal in which it is published. One legal document identification mechanism is Uniform Resource Name Namespace (URN) for Sources of Law (LEX) [23]. The identifier is conceived in such manner to depend only on the document characteristics and is independent of the document availability, access mode and physical location [23]. It has to be: globally unique, transparent, persistent, location-independent and language-neutral. Those characteristics provide mechanisms of stable cross-country references. Another legal document identification mechanism is specified in Architecture for Knowledge-Oriented Management of African Normative Texts using Open Standards and Ontologies (AKOMA NTOSO) [24]. The syntax of AKOMA NTOSO identifier is based on Uniform Resource Location (URL) standard [25]. With both mechanisms, it is possible to differentiate the identifiers of the level of FRBR work, expression, manifestation and item [26].

A legal act can be classified according to its subject matter. In Serbian legislature, there are eighteen classes of a legal act according its subject matter. Some of them are: defense, military and internal affairs; justice, criminal law and proceedings; trade, tourism and hospitality; public institutions, science, education, culture, media and sport; and many other. All of these classes have their subclasses. For example, trade, tourism and hospitality has four subclasses: trade, procurement and consumer protection; tourism and hospitality; protection of competition and state aid control; and stockpiles. The number of subclasses goes upto 23 [27].

After initiation phase, the document enters the preparation phase, i.e. the bill is drafted. As the content of the bill is strictly structured, it is possible to distinguish more than a few elements of the structure hierarchy: part, chapter, section, subsection, article, item, point, subpoint and line [25]. In our case, only articles and items appear as individuals of *DocumentFragment* class. It is important to notice that the model does not indicate whether the fragment is an article or an item. That can be infered from the name of the individual or by its place in the hierarchy.

The establishment phase it the phase in which the document, in this case, the bill, get enacted into law by the parliament, promulgated by the president, and published in the official gazette.

The use phase of a law in Serbia begins eight days after the law has been published at which point the law has legal consequences.

Document maintenance is carried out in the revision phase by enacting changes to existing legislation. As an example, we presented two versions of the law. These

versions are sequentially effective. In most cases, different versions of laws are sequentially effective, but there are some situations when there are two concurrently effective version of a law.

The law is repealed in the withdrawal phase.

V. CONCLUSION

In this paper, we presented some of the problems faced during design and development of DMSs and proposed a solution that is based on semantic web technologies. Documents were modeled starting from the concepts defined in the ISO 82045 family of the standards and time and provenance related concepts imported from Time Ontology and PROV-O.

Nevertheless, some problems still remain to be solved. There is a need to merge the presented ontology with metadata and classification ontologies. Furthermore, the static conceptualization of the document management domain has to be rethought in the context of business processes and workflow management. Enriching the model with Friend of a Friend (FOAF) is considered. FOAF is well-known ontology describing people and their relationships which can be used to improve security data. Also, our main goal, i.e. to customize the model to domains other than law and to implement components of the DMS described in [6], still has to be achived.

REFERENCES

[1] F. Castillo-Barrera, H. Durán-Limón, C. Médina-Ramírez, B. Rodriguez-Rocha, "*A method for building ontology-based electronic document management system for quality standards - the case of the ISO/TS 16949:2002 automotive standard*", Applied Intelligence, vol. 38, pp. 99-113, 2013

[2] International Organization for Standardization (ISO), "*ISO IEC 82045-1: Document Management – Part 1: Principles and Methods*," ISO, Geneva, Switzerland, 2001

[3] H. Zantout, F. Marir, "*Document Management Systems from current capabilities towards intelligent information retrieval: an overview*", International Journal of Information Management, vol. 19, Issue 6, pp. 471-484, 1999.

[4] A. Azad, "*Implementing Electronic Document and Record Management Systems*", chapter 14, Auerbach Publications, ISBN-10: 084938059, 2007

[5] T. Berners-Lee, J. Hendler and O. Lassila, "*The Semantic Web*", Scientific American, 2008.

[6] S. Gostojic, G. Sladic, B. Milosavljevic, M. Zaric and Z. Konjovic, "*Semantic Driven Document and Workflow Management*", International Conference on Applied Internet and Information Technologies (AIIT), 2014

[7] OWL DL [online], Avaliable at: http://www.w3.org/TR/2004/REC-owl-semantics-20040210/rdfs.html#5.4 [accessed January 14, 2015]

[8] MAchine Readable Cataloging (MARC) [online], Available at: http://www.loc.gov/marc/ [accessed January 14, 2015]

[9] Digital Object Identifier (DOI) [online], Available at: http://www.doi.org/ [accessed January 14, 2015]

[10] CrossRef [online], Available at: http://www.crossref.org/ [Accessed 20 Dec. 2014]

[11] DataCite [online], Available at: https://www.datacite.org/about-datacite/what-do-we-do [accessed January 14, 2015]

[12] R. Guenther, S. McCallum, "*New Metadata Standards for Digital Resources: MODS and METS*", Bulletin of the American Society for Information Science and Technology, pp12-15, ISSN: 0095-4403, 2003

[13] The DLM Foundation, "MoReq2010: Modular Requirements for Record Systems - Volume 1: Core Services & Plug-in Modules", [online], Available at: http://moreq2010.eu/ [accessed January 14, 2015]

[14] H.L. Kim, H.G. Kim, K.M. Park, "Ontalk: ontology-based personal document management system", WWW Alt. '04, May 2004

[15] Time ontology [online], Available at: http://www.w3.org/TR/owl-time/ [accessed January 14, 2015]

[16] PROV-O ontology [online], Available at: http://www.w3.org/TR/prov-o/ [accessed January 14, 2015]

[17] OWL 2 Web Ontology Language Manchester Syntax (Second Edition) [online], Available at: http://www.w3.org/TR/owl2-manchester-syntax/ [accessed January 14, 2015]

[18] Dublin Core Metadata Initiative (DCMI) [online], Available at: http://dublincore.org/specifications/[accessed January 14, 2015]

[19] C. McGregor, O. Alonso, S. Alpha, S. Buxton et al., "*Oracle Text Application Developer's Guide, 10g, Release 1 (10.1)*", chapter 6 "*Document Classification*"

[20] PROV-Overview [online], Available at: http://www.w3.org/TR/2013/NOTE-prov-overview-20130430/[accessed January 14, 2015]

[21] Document Management Ontology [online], Available at: http://www.informatika.ftn.uns.ac.rs/82045-1 [accessed January 14, 2015]

[22] Narodna Skupština Republike Srbije, "*Law on Personal Data Protection*", Službeni glasnik Republike Srbije no. 104/2009

[23] P. Spinosa, E. Francesconi, C. Lupo, "*A uniform resource name (URN) namespace for sources of law (LEX)*", Internet Engineering Task Force, Fremont, 2011, available at: http://tools.ietf.org/html/draft-spinosa-urn-lex-04 [Accessed February 20, 2015]

[24] Akoma Ntoso [online], Available at: http://www.akomantoso.org/, [accessed February 20, 2015]

[25] S. Gostojic, "*Kreiranje i korišćenje digitalnih dokumenata pravne regulative*", Doctoral dissertation, University of Novi Sad, 2012

[26] International Federation of Library Associations and Institutions, "*Functional Requirements for Bibliographic Records*", International Federation of Library Associations and Institutions, The Hague, 2007, available at: http://www.ifla.org/en/publications/functionalrequirements-for-bibliographic-records [Accessed February 20, 2015]

[27] Narodna Skupština Republike Srbije, "*Constitution of the Republic of Serbia*", Službeni glasnik Republike Srbije no. 98/2006

SilabMDD - A Use Case Model Driven Approach

Dušan Savić, Siniša Vlajić, Saša Lazarević, Vojislav Stanojević, Ilija Antović, Miloš Milić[1], Alberto Rodrigues da Silva[2]

Faculty of Organizational Sciences, University of Belgrade [1]
Department of Computer Science and Engineering
IST / University of Lisbon[2]

Abstract - *Model-Driven Development (MDD) is a software development paradigm that emphasizes the importance of using models during the entire software development process, models with different levels of abstraction. In our SilabMDD approach, use cases models allow to define user and software requirements. These models are specified in the SilabReq language which is implemented inside JetBrains Meta Programming System (MPS) and can be used as plug-in for IntelliJ IDEA or for the MPS tools.*

1. INTRODUCTION

The development of information systems is a complex and social process because it involves many interactions among different stakeholders. In order to make this process successful it is necessary to understand the system requirements and document them in a suitable manner. There are multiple definitions for requirements, namely: (1) a property that must be exhibited in order to solve some real-world problem [1]; (2) needs and constraints placed on a software product that contribute to the solution of some real-world problem [2]; (3) a condition or capability needed by a user to solve a problem or achieve an objective; or (4) a condition or capability that must be met or processed by a system (or system component) to satisfy a contract, standard, specification, or other formally imposed documents [3]. Additionally, there are many forms for requirements presentation such as natural language, constrained natural language or model based requirements language [4].

Requirements engineering (RE) involves two main processes [4]: (1) requirements development (with elicit, analyze, specify, and validate software requirements) and (2) requirements management process. Many RE approaches have been discussed in the literature, which differ in their methods, modeling techniques and modeling languages. Widely accepted approaches in 70ies and 80ies were mainly data and functional-oriented analysis techniques, while object-oriented approaches were emerged and were popular during the late 80s and 90s. Another approach emerged more recently were goal-oriented requirements engineering [5]. Common to all these approaches, particularly in their early period, is the clear separation of RE process from software development process.

Other software paradigm, referred to as Model-Driven Development (MDD) [6], is a software paradigm that emphasizes the importance of models. The aim of MDD is to use models throughout the software development process at different levels of abstraction. Therefore, models are not used only to document some part of a system; models are first-citizen in software development. MDD processes usually start to develop a requirements model which is defined to describe user's needs in a computational independent way. Then, this model can be refined into one or more models that describe the system without considering technological aspects. Finally, these models are either refined into design models (that describe the system by using concepts of a specific technology) and are then translated into a source code; or are directly derived to a code if they contain enough information to implement the software system in a precise and complete way [7].

However, despite the importance of RE as a key success factor for software development projects there is still a lack of MDD methods that would cover the full development lifecycle, from a RE level to a development level with source code generation or writing activities [8] [9].

In this paper we introduce SilabMDD approach that is a use case and MDD approach that use SilabReq as a use cases specification language. Furthermore, we present how SilabReq is supported by JetBrains Meta Programing System (MPS). The goal of SilabMDD is to provide a complete software development workbench to be used by requirements engineers, developers, as well as by non-technical stakeholders.

This paper is organized as follows. Section 2 describes the background of this work. Section 3 presents SilabMDD approach while Section 4 concludes the paper and outlines future work.

2. BACKGROUND

Requirements are mostly documented using natural language. However, natural language requirements specification tends to be ambiguous, unclear, and inconsistent [4]. On the other hand, documenting requirements using semi-formal models require using specific modeling language for each particular perspective. Pohl proposed three types of requirements proposed by [4]: (1) goals, which document intentions of stakeholders; (2) scenarios, that describe concrete example of system usage; and (3) solution-oriented requirements, that can be used as complement each other. Different requirements modeling languages can be used for modeling different requirements artifacts for different perspective. For example, i* [4] and KAOS[4] goal oriented models can be used for specification of the goals; UML use case, sequence and activity diagram can be used for modeling scenarios; while UML state machines or data flow daigrams can be used for modeling system's behavior.

The specification of requirements is a difficult task because requirements are read by many participants of the software development process with different technical knowledge. People prefer to use textual specification of requirements, but their representations are not suitable for automatic validation, transformation and even reusing. We need a structured language for requirements specification that will be understandable by most of these participants but also will be precise enough to enable automatic validation and transformation. This language should be defined by meta-modeling or grammar ware in order to enable automatic or semi-automatic processing.

UML is a standard language for modeling software systems and many people have also used it for requirements specification. However, some authors have argued that UML has some deficiencies as a semiformal requirements specification language [10].

There are other Requirements Specification Language (RSL) that use natural language in a controlled way. Smiałek at el. defined RSL as a semiformal natural language that employs use case for specifying requirements [11]. Each scenario in a use case contains special controlled natural language SVO (O) sentence. RSL has been developed as a part of ReDSeeDS project [12]. ReDSeeDS approach covers a complete chain of model-driven development – from requirements to code[13] .

The goal of ProjectIT [14] [15] is to provide a complete software development workbench, with support for project management, requirements engineering, analysis, design and code generation activities. ProjectIT-Requirements is the component of the ProjectIT architecture that deals with requirements issues. The main goal of the ProjectIT-Requirements is to develop a model for the definition and documentation of requirements, which, by raising their specification rigor, facilitates the reuse and faster the integration with MDD development environments driven by models. Taking into account the different types of requirements, this project uses software requirements, those that can more easily be “converted” in software design models by MDD approaches [16] .

Recently, the ProjectIT's RSL evolved to a more flexible approach named RSLingo[17][18]. RSLingo is a linguistic approach for improving the quality of requirements specification, which is based on two languages and mapping between them: the RSL-PL and the RSL-IL. RSL-PL (Pattern Language) is an extensible language for defining linguistic patterns dealing with information extraction from requirements written in natural language. On the other hand, RSL-IL (Intermediate Language) is a formal language with a fixed set of constructs for representing and conveying RE-specific concerns.

3. THE SILABMDD APPROACH

SilabMDD approach emerged as a key result of Silab Project which was initiated in 2007 in the Software Engineering Laboratory at Faculty of Organizational Sciences, University of Belgrade. The main goal of this project was to enable automated analysis and processing of software requirements in order to achieve automatic generation of different parts of a software system. In the beginning, Silab Project has been divided in two main sub-projects SilabReq and SilabUI that were being developed separately. Initially the SilabReq project focused on the formalization of user requirements and their transformations to different UML models to facilitate the analyses process and to assure the quality of software requirements. On the other hand, SilabUI project focused on automatic generation of user interfaces from use cases specifications. When both subprojects reached the desired level of maturity, they were integrated in a way that the results of SilabReq were used as input for SilabUI project. As a proof of concept, Silab project was used for the Kostmod 4.0 [19] project, which was implemented for the needs of the Royal Norwegian Ministry of Defense.

After several years of using this project in developing different intensive software system we are established a SilabMDD approach. This section introduces the conceptual view of the SilabMDD approach.

A. Overview of the SilabMDD approach

Fig. 1 depicts the key artifacts of SilabMDD approach. SilabMDD approach is a use case and model driven approach.

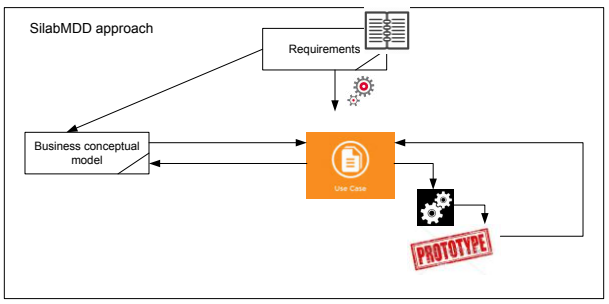


Figure 1. Modeling circle in SilabMDD approach

Usually in MDD the source code is (semi) automatically generated from the models. Despite the fact that use cases are narratives, there is no single standard that specify what textual specification of use case should be. SilabMDD approach includes SilabReq language that is a DSL used for use case specification. SilabReq language allows requires a rigorous definition of the use case specification, particularly description of sequences of action steps, pre- and post-conditions, and relationships between use case and elements (classes) defined in domain models.

SilabMDD approach is also language-oriented. Language Oriented Programming [20] presents a style of development which operates about the idea of building software around a set of DSL, while Language

Workbench is a generic term for tools that use this style of programming. Meta Programming System[1] (MPS) from JetBrains[2] is one of the most popular meta-programming tool that enables language oriented programming with a projection editor in persistent abstract representation that was used to develop SilabMDD's languages.

SilabMDD is use-case driven approach but it do not pay much attention to the way in which use-cases are elicited. They can be derived from business process or from text requirements. If requirements are expressed in some form of model as in RSLingo (using RSL-IL[17]) it is possible to automatically use appropriate transformation to deliver use cases. Integrated in this the specification process, use cases can be specified using SilabReqUC language and continuous inspection of business model. For the description business conceptual model we propose the SilabReqBCM language. Use case actions, pre-conditions and post-conditions are specified in the context of business conceptual models. Therefore, SilabMDD approach use SilabReqUI language which is primary use for specification user interface prototype.

B. *Specification of use case from different level of abstraction*

The SilabReq allows defining use cases specifications at different levels of abstractions according to the different roles involved on this process . There are three different abstraction levels: (1) *use case interaction level* (high-level), (2) *use case behavior level* (medium-level), and (3) *use case UI-based level* (lower-level) [21][22]. Each abstraction level extends and semantically enriches the previous level. Actually, we can use the same model for both user and system requirements. Transformations among these different levels are used to create multiple views as well as for code generation.

Use case action can divide in two categories: (1) the actions performed by the users and (2) the actions performed by system [21,22]. Both categories contain different types of actions. In the category in which actions are performed by the user, we have identified actions types such as: (1.1) Actor Prepare Data to execute System Operation (APDExecuteSO) and (1.2) Actor Calls System to execute System Operation (ACSExecuteSO). On the other hand, in the category in which actions are performed by the system, we have identified two action types such as: (2.1) System executes System Operation (SExecuteSO) and (2.2) System replies and returns the Result of the System Operation execution (SRExecutionSO).

The main task that use cases have *at the highest level of abstraction (interaction level) is user requirements specification.* Therefore, this level allows non-technical stakeholders to quickly read and understand use case descriptions. Use cases alone are not sufficient for user requirements specification. Therefore, use cases are complemented with glossary and business rules. Glossary and business rules are specified within the same language (but it is possible to create and use other specific languages). They are specified separately, but connected with some elements of use cases such as pre-conditions, post-conditions or use case actions. Business rule and terms (in the glossary) are specified with unique identification, name and description. At this level of abstraction, each use case specification consists in the following elements (see Fig.2): unique use case identifier, use case name, the actors who participate in use case, the business entity over which the use case is executed, main and alternative use case scenarios, and use case pre-conditions and post-conditions.

Business rules are used for specification of use case pre-conditions and post-conditions. *Pre-conditions* and *post-conditions* are specified in the context of the system state as a pair of entities and its state. In pre-condition, the system state defines the conditions that must be satisfied before use case starts. This business rules are specified in context of business domain model. Therefore, before use case starts, business entity over which the use case is executed and related entity must be in some specific state. For example, the user must be logged in as administrator (user as an entity and login as a state), the order must exist (order as entity, exist as state), the form for creating bill is open and the order exist (form for bill as state and open as state, order as entity and exist as state). The similar situation is applied to post-conditions specification. After successful execution of a use case, its entity of the system will be in some particular state (for example order as entity will be saved). Action business rules are used to specify data entered by actor (choose or select), or data returned from system. At this level of abstraction, these rules are related with APDExecuteSO action and SRExecutionSO action. Both of these actions are specified in context of business domain model. The Fig. 2 describes the use case specification template document, entity and business rule specification document. The specification document looks like a wiki based document in the way that it is possible to navigate through document.

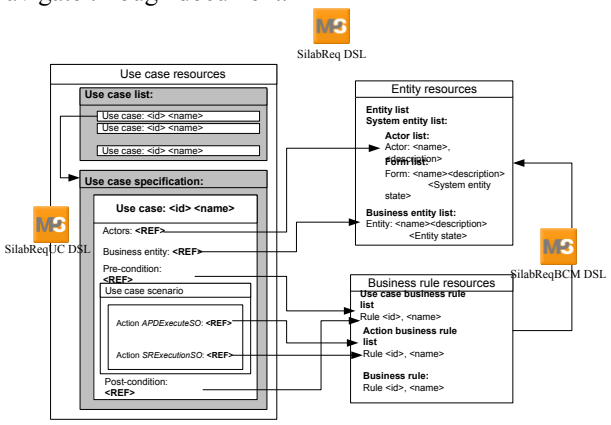


Figure 2 Use case specification from user perspective

The specification of use cases at the medium level is used to determine the desired functionality of the system. At this level, a use case scenario specification is extended with the specification of ACSExecuteSO and SExecuteSO

[1] http://www.jetbrains.com/mps/
[2] http://www.jetbrains.com/

actions. These use case actions are used to define a function that a system should provide. Fig.3 describes how medium-level use case specification extends the high-level specification. This figure describes part of previous template document, which is extended with the specification system operations that contains: system operation pre-condition, successful and error system response, as well as system operation post-condition.

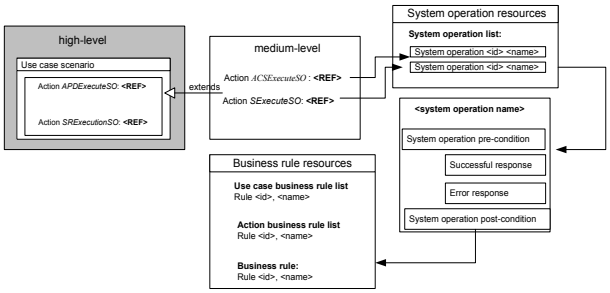


Figure 3 Use case specification from behavior perspective

We introduce ACSExecuteSO action because the user can call system operation in different ways (for example double click on button, pressing specific key on keyboard, focusing on some graphic user component and etc.). At this level, we just emphasize that user just define system operations, but the way how the user does it is only specified at the lowest level. As a result of this level of use case specification, we have identified and specified system function as system operation contracts.

The lower-level of use case specification includes the details about specification user interface. This specification is done in several steps. First, we define appropriate template (e.g. filed-form, table-form, filed-tab, table-tab) which is used to display the main business entity and entities associated with it. Second, from the use case business rule we identify entity and entity attributes that participates in use case and specify the corresponding graphic user interface component used to display and modify its value (e.g. text field, table, dropdown list, radio buttons).Third, for each ACSExecuteSO action we specify the graphic user interface components (e.g. button, menu item) that are used to call system to execute the system operation. Fig.4 suggests the relations between use cases, use case templates, business rules and business entities with corresponding GUI component.

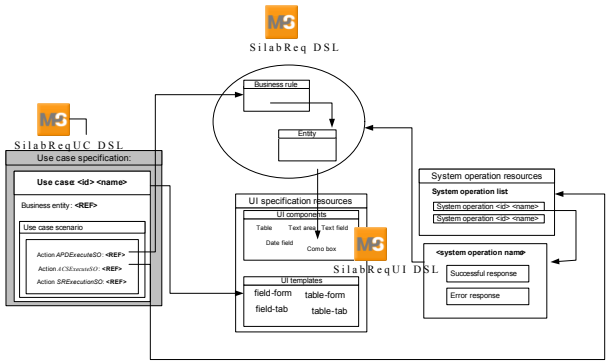


Figure 4 Use case specification from UI perspective

4. THE SILABMDD TOOL SUPPORT

The SilabMDD approach is supported by companion tool, also named SilabMDD tool, that has been developed on top of JetBrains MPS. MPS contains its own language named BaseLanguage. MPS allows extending BaseLanguage to create new custom languages, extend existing languages, and use them to develop software applications. BaseLanguage has a built-in support with strings, collections, regular expressions, etc. During the process of creating a new language it is needed to derive concepts from the BaseLanguage as a reference for new languages.

The major goal of MPS is to allow languages definitions thought extension, which means that language's designer, can use concepts from a new extended language as well as combine concepts from different languages. The problem in syntax language extension is mainly the textual concrete syntax because each language may have its own concrete syntax. JetBrains MPS proposes having concrete syntax maintained in an Abstract Syntax Tree (that consists of nodes with properties, children and references that describes the program code). At the same time, MPS offers an efficient way to keep writing code in a text-like manner.

MPS uses a generative approach that focuses on automating the creation of system-family members: a given system can be automatically generated from a specification written in one or more textual or graphical domain-specific languages [23].

From a developer perspective, this programming approach seems very promising: developers have two ways to implement software. First, they can use requirements specification to manually and formally define their requirements. Second, they can use or create different transformation to generate source code from these models. Both alternatives can be used and integrated with MPS because there is already a plug-in for IntelliJ IDEA which allows including MPS concept models in Java project. So, developers can use MPS for writing Java application and integrate Java source code with SilabReq requirements specifications.

MPS comes with sets of DSL which is use to define the structure of language, editor, type systems, and generators. All of these DSLs are built using MPS itself. The language definition starts by defining its abstract syntax us suggested in Fig.5 (concepts in MPS). The concept is one element of language in MPS which describes how the elements look like, behave and generate[3]. Each concept can have a definition in one or more aspects of language such as structure, editor or generator.

[3] http://dslbook.squarespace.com

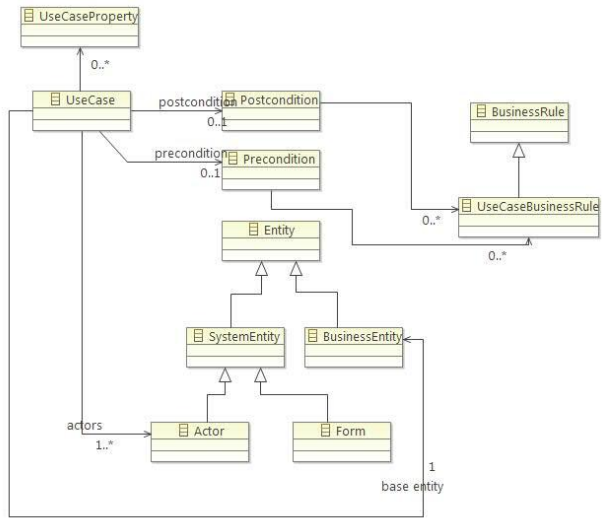


Figure 5 The SilabReq meta-model (partial view)

This part of SilabReqUseCase specification language is described in MPS using its Concept Declaration Editor (Fig.6).

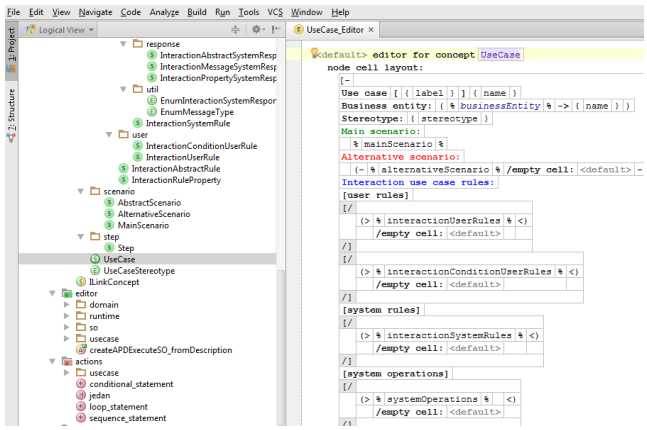


Figure 6 MPS editor for concept declaration

The MPS' Aspect Editor is used for defining the concepts' for concrete syntax.Fig.7 depicts the using of Aspect Editor for UseCase concept.

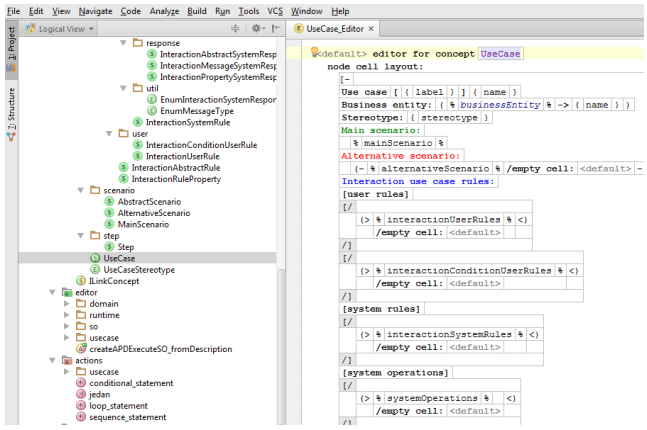


Figure 7 Aspect Editor for Use Case concept

We use MPS to generate Java source code from requirements specification model. Java is embedded into MPS, so generation Java source code is a simple transformation. We use template – based transformation in MPS for generation source code. This transformation has two main important building blocks: mapping rules (define which concepts are processed with which templates), reduction rules (define transformations which removes source node and replace it with associated template) and templates. Fig.8 presents an example of transformation declaration in MPS.

```
mapping labels:
  label inputFieldDeclaration : InputFiled        -> FieldDeclaration
  label useCaseTemplate       : UIUseCaseTemplate -> TemplateGeneration

parameters:
  << ... >>

is applicable:
  <always>

conditional root rules:
  << ... >>

root mapping rules:
  concept          UIUseCase  --> UIUseCase
  inheritors       false
  condition        <always>
  keep input root default
```

Figure 8 Example of MPS transformation definition

5. CONCLUSION

In this paper we introduce a SilabMDD use-case and model driven approach that includes the SilabReq, a use case specification language. Further, we present how SilabReq is supported by JetBrains Meta Programing System (MPS) framework. SilabMDD is a use case driven approach but it do not pay much attention to the way are elicited use cases. It requires a rigorous definition of the use case specification, particularly description of sequences of action, pre- and post-conditions, and relationships between use cases and business entities.

The goal of SilabMDD is to provide a complete software development workbench to be used by requirements engineers, developers, as well as by other non-technical stakeholders.

In short, the contributions of this article are twofold. Firstly, it introduce SilabReq specification language which can be used for requirements specification. Secondary, it presents SilabMDD as use-case and model driven approach. In this approach use case become the key and central artifact in software development process which are considered at different levels of abstraction.

REFERENCES

[1] IEEE Computer Society Professional Practices Committee SWEBOK®, Guide to the Software Engineering Body of Knowledge. The Institute of Electrical and Electronics Engineers, Inc., 2004

[2] G.Kotonya and I. Sommerville, Requirements Engineering Processes and Techniques. John Wiley and

Sons, 2000Banks, J. and S. J. Carson, Discrete-Event System Simulation, Prentice-Hall, New Jersey, 1984.

[3] IEEE standard glossary of software engineering terminology,IEEE Std610.12-1990, 1990

[4] K.Pohl, Requirements Engineering - Fundamentals, Principles, and Techniques. Springer 2010

[5] A. van Lamsweerde, "Goal-Oriented Requirements Engineering: A Roundtrip from Research to Practice," Requirements Engineering, vol. 6, no. 11, pp. 4–7, 2004.

[6] S. Mellor, A. Clark and T. Futagami, "Model-Driven Development," IEEE Software, vol. 20, pp. 14-18, 2003.

[7] P. Valderas and V.Pelechano, "A Survey of Requirements Specification in Model-Driven Development of Web Applications", TWEB 5(2):10 (2011)

[8] T.Menzies, ″Editorial: model-based requirements engineering″, Requirements Engi 8(4): 193-194, 2003

[9] G.Loniewski, E. Insfran and S. Abrahão, "A systematic review of the use of requirements engineering techniques in model-driven development", Model driven engineering languages and systems. D. Petriu, N. Rouquette and Ø. Haugen (ed.), Springer: 213-227, 2010

[10] M.Glinz, "Problems and Deficiencies of UML as a Requirements Specification Language", Proc. of the 10th IEEE Int. Workshop on Software Specification and Design, 2000

[11] M. Smiałek, J.Bojarski, W.Nowakowski and T. Straszak, "Scenario construction tool based on extended UML metamodel". Lecture Notes in Computer Science, 3713:414–429, 2005.

[12] M.Smialek and T.Straszak, "Facilitating transition from requirements to code with the ReDSeeDS tool". RE 2012: 321-322

[13] M.Smialek, W.Nowakowski, N. Jarzebowski, A Ambroziewicz," From use cases and their relationships to code" MoDRE 2012: 9-18

[14] A.Silva, C.Videira, J.Saraiva, D.Ferreira and R.Silva, "The ProjectIT-Studio, an integrated environment for the development of information systems", In Proc. of the 2nd Int. Conference of Innovative Views of .NET Technologies (IVNET'06), pages 85–103. Sociedade Brasileira de Computação and Microsoft.

[15] A. R. d. Silva, J. Saraiva, D. Ferreira, R. Silva, and C. Videira, "Integration of RE and MDE Paradigms: The ProjectIT Approach and Tools", IET Software: On the Interplay of .NET and Contemporary Development Techniques, 2007

[16] D. A. Ferreira and A.R.Silva,"A Controlled Natural Language Approach for Integrating Requirements and Model-Driven Engineering", ICSEA 2009: 518-523

[17] D. A. Ferreira and A.R.Silva,"RSLingo: An information extraction approach toward formal requirements specifications", MoDRE 2012: 39-48

[18] Alberto Rodrigues da Silva, João Saraiva, David Ferreira, Rui Silva, Carlos Videira, Integration of RE and MDE Paradigms: The ProjectIT Approach and Tools, in IET Software Journal - Special issue "On the interplay of .NET and contemporary software engineering techniques", December 2007, Volume 1, Issue 6, p. 217-314, IET.

[19] Kostmod4.0 http://rapporter.ffi.no/rapporter/2009/01002.pdf, accessed in January, 2013

[20] F, Martin. Language Workbenches: The Killer-App for Domain Specifc Languagess [online]. Available on: http://martinfowler.com/articles/languageWorkbench.html

[21] D.Savić, A.Rodrigues da Silva, S.Vlajić, S.Lazarević, I.Antović, V. Stanojević, M.Milić, Preliminary experience using JetBrains MPS to implement a requirements specification language, in Proceedings of QUATIC'2014 Conference, 2014 , IEEE Computer Society .

[22] D.Savić, A.Rodrigues da Silva, S.Vlajić, S.Lazarević, I.Antović, V. Stanojević, M.Milić, Use Case Specification at Different Levels of Abstraction, in Proceedings of QUATIC'2012 Conference, 2012 , IEEE Computer Society.

[23] K.Czarnecki, Generative Programming: Methods, Tools, and Applications. Addison-Wesley (2000)

Service Networks Monitoring for better Quality of Service

Tehreem Masood, Chantal Bonner Cherifi, Néjib Moalla
University of Lyon 2, DISP Laboratory, Lyon, France
Tehreem.Masood@univ-lyon2.fr, chantal.bonnercherifi@univ-lyon2.fr, Nejib.Moalla@univ-lyon2.fr

Abstract— ***Today, the deployment of Web services in many enterprise applications has gained much attention. Service network inhibits certain common properties as they arise spontaneously and are subject to high fluctuation. The objective of consumer is to compose services for stable business processes in coherence with their legacy system capabilities and with better quality of services. For this purpose we have proposed a dynamic decision model that integrates several performance metrics and attributes to monitor the performance of service oriented systems in order to ensure their sustainability. Based on the available metrics, we have identified performance metrics criteria and classified into categories like time based QoS, size based QoS, combined QoS and estimated attributes. Then we have designed service network monitoring ontology (SNM). Our decision model will take user query and SNM as input, measures the performance capabilities and suggests some new performance configurations like selected service is not available, physical resource is not available and no maintenance will be available for the selected service for composition.***

Keywords: Web Services, SOAP, Service-Oriented Architecture, Monitoring, Simulation, Quality of Service parameters, Performance, Decision model

INTRODUCTION

Service-Oriented Computing (SOC) increasingly gains motivation in both industry and academia as a means to develop adaptive distributed software applications in a loosely coupled way. The motivation behind SOC is the idea that businesses offer their application functionality as services over the Internet and other users or companies can integrate and compose these business services into their applications [1]. Web services became very important during the past few years. It is a software system designed to support interoperable interaction between different applications and different platforms [2]. Web services use standards such as Hypertext Transfer Protocol (HTTP), Simple Object Access Protocol (SOAP) [3], Universal Description, Discovery, and Integration (UDDI) Web Services Description Language (WSDL) [4] and Extensible Markup Language (XML) for the communication between web services through internet.

Web services flow specifications like business process execution language for Web services (BPEL4WS) [5] and Web services choreography interface (WSCI) have also been discussed in literature [5]. Since business requirements are becoming the major driving force for creating Web services research topics to support the business process integration, collaboration, and management, the business context should be captured and transmitted into appropriate partners [5].

There are two levels of performance problems of Web services, namely system level which is related to SOAP and XML, and server level which is related to the processing of SOAP requests at the server side [6]. If we deal with them properly, it will definitely provide a more efficient and scalable structure in terms of performance for deploying and running Web services. Web services are supposed to be a source of generating increased returns for enterprises by exposing the legacy enterprise applications to a wide range of other applications on different platforms [7]. It is important to address the better quality of service technique that makes efficient use of available resources.

Business service developers are just to assemble a set of appropriate web services to implement the business tasks. Business applications are no longer written manually. For example [8], a client requirement can be expressed as a sequence diagram in UML. It is composed of several sub-functional modules or abstract services. Each abstract service is associated with a web service community which contains several existing web services with the same functionality. The process of selecting a service from a web service community for an abstract service by quality of service attributes is called the local selection. As a task presented by the service composition can be explained by a significant number of combinations.

The objective of consumer is to compose services for the business collaboration. Constraints are to not impact the performance of legacy systems as well as the service composition for stable business and with better quality of services. For this purpose we have proposed a dynamic decision model that integrates several performance metrics and attributes to monitor the performance of service oriented systems in order to ensure their sustainability. Based on the available metrics, we have identified performance metrics criteria and classified into categories. Then we have designed service network monitoring ontology (SNM). Our decision model will take user query and SNM as input, measures the performance capabilities and suggests some new performance configurations like selected service is not available, physical resource is not available and no maintenance will be available for the selected service for composition.

The remaining paper is organized as follows: Section II includes related work in the area of performance of web services. Section III discusses the research challenges. Section IV discusses the proposed approach. In the end, conclusion of the paper is presented in Section V.

RELATED WORKS

In this section we have classified the related work into three categories. System level performance, server level performance and different web services architectures. These three types of techniques will help reader to understand about the existing work of different performance levels like domain, node, service, server and messaging. We are concerned with all these performance levels.

1. *System Level Performance*

The System Level covers domain level, messaging or service level, node level and service level performance. In this category we have discussed some techniques that are related to system level performance.

Z. Tari et al. [10] proposed a benchmark of different SOAP bindings in wireless environments. Its configuration and results can serve as a standard benchmark for other researchers who are also interested in the performance of SOAP bindings in wireless networks. Three sets of experiments were carried out: loopback mode, wireless network mode and mobile device mode. The experimental results show that HTTP binding inherits very high protocol overhead (30%–50% higher than UDP binding) from TCP due to the slow connection establishments and tear-down processes and the packet acknowledgement mechanism. UDP binding has the lower overhead because it does not require establishing connections before transmitting datagrams and does not address reliability. This results in a reduction in the response time and an increase in the total throughput.

Z. Tari et al. [11] proposed a similarity-based SOAP multicast protocol (SMP) which reduces the network load by reducing the total generated traffic size. It is based on the syntactic similarity of SOAP messages. In particular, the SMP reuses common templates and payload values among the SOAP messages and only sends one copy of the common part to multiple clients. SMP makes use of the commonly available WSDL description of a SOAP Web service when determining the similarity of response messages. For messages that are highly similar, instead of generating messages with duplicated similar parts for different clients, the duplicated parts are reused for multiple clients and are sent only once from the source .

Z. Tari et al. [12] proposed a tc-SMP1 technique which is an extension of SMP, which is the traffic-constrained similarity-based SOAP multicast protocol (tc- SMP) is proposed here. Two algorithms, greedy and incremental approaches, are described to address this problem. Both tc-SMP algorithms aim at minimizing the total network traffic of the whole routing tree every time a new client is added to the tree. Two heuristic methods are also proposed for these algorithms to assist in choosing the order of clients being added to the tree. In general, the performance improvement of tc-SMP is about 30% higher network traffic reduction than SMP at a small expense of up to 10% rise in the response time.

Comparison of the system level performance techniques is described in Table I. Parameters used in the comparison table are response time, throughput, network traffic, binding type and similarity matching.

Response time:

It is also called latency. It is the time perceived by a client to obtain a reply for a request for a web service. It includes the transmission delays on the communication link. It is measured in time units

Throughput:

The number of requests executed per unit of time. For web service users it can be measured by requests per seconds or number of operations per second

Network Traffic:

The total network traffic for communication scheme or session which is the number of bytes transferred during the communication

Binding Type:

Binding type is the type of protocol used for binding. Either it is http, udp or more

Similarity Matching:

Similarity matching is the type of matching like syntactic or semantic. NA in Table 1 means that this parameter is not applicable to the corresponding technique

TABLE I
COMPARISON OF SYSTEM LEVEL PERFORMANCE TECHNIQUES

Techniques		Parameters				
		Response Time	Throughput	Network Traffic	Similarity Matching	Binding Type
SOAP Binding	SOAP over HTTP	Large	Small	Large	HTTP	NA
	SOAP over TCP	Medium	Medium	Large	TCP	NA
	SOAP over UDP	Small	Large	Small	UDP	NA
The Use of Similarity & Multicast Protocols to Improve Performance SMP [7]		Medium	Medium	Medium	SOAP Over HTTP	Syntactic
Network Traffic Optimization tc-SMP1 [8]		Large	Large	Small	SOAP Over HTTP	Syntactic

2. *Server Level Performance*

In this category we have discussed some techniques that are available for server level performance.

A series of advanced task assignment policies: TAGS (Task Assignment by Guessing Size), TAGS-WC (TAGS with Work Conserving), TAPTF (Task Assignment based on Prioritizing Traffic Flows), TAPTF-WC (TAPTF with Work Conserving), and MTTMELL (Multi-Tier Task Assignment with Minimum Excess Load) Policy.

Multi-level Time Sharing (MLTP) Policy investigated time sharing under more general setting where amount of time service time (quantum) allocated on levels using a random variable.

3. *Different Web Services Architectures*

In this category, we have discussed various web services architectures that cover different level of performance like messaging or service level and node level performance.

Zhou et al. [9] proposed UX which is an extended UDDI. It assesses previous and current service usage for the future service selection. With the analysis of the network model, the condition of service requester's connection is recorded by the server to enable better predictions in a future service's request. Instead of the QoS description published by service provider, QoS feedbacks made by service requesters are used to generate summaries for the invoked services. These summaries are then used to predict the services' future performance. A general federated service is designed so that server nodes can be administratively federated across network boundaries. Based on this federated service, lookup interface is provided on a UX server that facilitates the discovery between different registries and the exchange of service QoS summaries.

Bertino et al. [13] proposed an approach based on Merkle hash trees, which provides a flexible authentication mechanism for UDDI registries. They have claimed two relevant benefits. The first is the possibility for the service provider to ensure the authenticity and integrity of the whole data structures by signing a unique small amount of data, with the obvious improvement of the performance. The second benefit regards browse pattern inquiries is that they return the overview information taken from one or more data structures. According to the UDDI specification, in such a case if a client wishes to verify the authenticity and integrity of the answer, it must request the whole data structures from which the information are taken.

Curran and Gallagher [14] proposed a framework called Webber, which provides the services necessary for supporting new communication protocols and qualities of service. Webber consists of a set of Java classes for representing the uniform resource locators, protocol stacks, the framework API and SOAP. The abstraction is analogous to that of the various broadcast media in everyday use, such as newspaper, radio and TV corresponding to text, audio and video components contained in multimedia applications. Webber fragments the various media elements of a multimedia application and 'broadcasts' them over separate channels to be subscribed to at the receiver's own choice.

Mateos et al. [15] proposed a scheme named as MoviLog. MoviLog is a platform for building the intelligent mobile agents, based on a strong mobility model, where agents' execution state is transferred during migration. MoviLog is an extension of JavaLog a framework for an agent-oriented programming. In order to provide mobility across sites, each MoviLog host has to execute a MARlet which is a mobile agent resource. A MARlet is a Java servlet that encapsulates a Prolog inference engine and provides services to access it. In this way, a MARlet represents an execution environment for mobile agents, or brainlets in MoviLog terminology. In this sense, a MARlet can be used as an inference server for agents and external Web applications. This mechanism states that when certain predicates previously declared in the code of a Brainlet fail, MoviLog transparently moves the Brainlet and its execution state to another site that contains definitions for that predicate, thus making local use of those definitions later.

Comparison of some of web services architectures level technologies of web services discussed in this paper is described in Table II. Parameters used in the comparison table are response time, reliability, standard used and cost. Response time has already been explained. Other parameters are explained below.

Reliability:

Reliability corresponds to the likelihood that the service will perform when the user demands it and it is a function of the failure rate. Each service has two distinct terminating states: One indicates that a Web service has failed or aborted, the other indicates that it is successful or committed

Standard Used:

It describes the type of standard used in the corresponding technique

Cost:

Cost represents the cost associated with the execution of the service.

TABLE II
COMPARISON OF SOME WEB SERVICE ARCHITECTURES

Techniques	Parameters			
	Reliability	Response Time	Standard	Cost
QoS-Aware Web Services(UX) [5]	Medium	Large	UDDI	High
Merkele [9]	High	Large	UDDI	High
Webber [10]	High	Large	SOAP	High
MoviLog [11]	High	Small	OWL S	High

III. RESEARCH CHALLENGES

There are several challenges that have been addressed in the literature in order to monitor the performance of web service network at different levels. All the levels are shown in Figure 1.

Data abstraction layer is used to query data from the database and the historical information about data and components which is available in the legacy. Data abstraction layer will provide the data services to the messaging layer. Once data services are gathered then messaging layer provides the ability to perform the necessary message transformation to connect the service requestor to the service provider and to publish and subscribe messages and events asynchronously. In this way services are published in the service layer. Then we have the whole service oriented architecture in the process orchestration layer. All this information is stored in the UDDI. The next layer of BP application provides the service composition. Governance rules are the set of policies like service will be available for one year etc.

Security is used to provide some integrity to the system like authentication with the help of user name and password. Service metrics are the parameters in order to guarantee the performance of web services. For example in the information technology infrastructure library (ITIL) there are 5 sub-categories and more than 100 metrics available for service support process and 5 sub-categories and more than 50 metrics available for service delivery processes. Operational intelligence and audit component is used to add some parameters in the query to measure the performance of the services. BP state intelligence is used to provide some intelligence or flags to measure the performance of BPEL. BAM (Business activity monitoring) and ARM (Application response time measurement) are the infrastructures that will help to measure different quality metrics.

This shows that different kinds of technologies and infrastructures are available that can be utilized to measure the performance of service network at different levels. But they are not effectively utilized in order to provide better quality of service in service oriented architectures (SOA). We can measure generic traces for web service network monitoring like daily, weekly and monthly trend in the value fluctuation, loss or error of service, un availability of service, SLA violations, resources are not available and new service requirement.

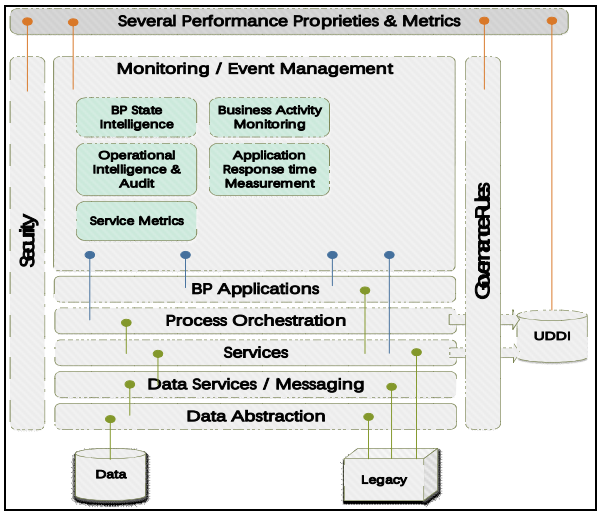


Figure 1. Research Challenges of Performance of SOA

IV. PROPOSED APPROACH

First of all, the identification of performance criteria from the available performance metrics has been performed. Several performance metrics are available for decision. In Figure 2, we have shown different levels of performance with the help of lending example. Customer wants to take loan from the loan organization. Loan organization is the domain that has many services like receive application, check credit, negotiate loan, close loan and book loan. At the node level we have shown the composition of services like customer accounting is the composition of check credit and book loan. Similarly credit administration is the composition of check credit and negotiate loan. Service messaging is related to the protocols used. Loan organization system is the server that manages the physical resources as well as receives and modifies application. We have shown all these performance levels in our ontology discussed in Section IV.

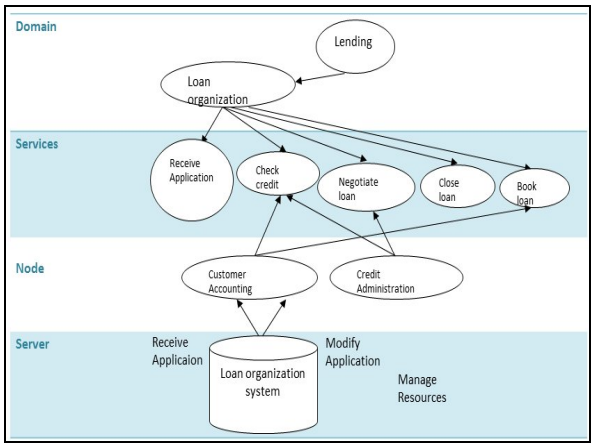


Figure 2. Example

The next step is the classification of these metrics and attributes based on some criteria. Qos metric values are gathered from the metrics database and used as attributes of a record of the Matching set. Then, the choice of KPI for the evaluation based on the user requirement by selecting KPI target value and analysis period in the structuring phase. The Predicate of KPI metric is classified as valid or violated. Then a decision model have been proposed that will generate answers like the selected service is not available for an additional consumption, no physical resources are available to support the new deployment, no maintenance will be available for the selected service for composition and security compliance problem in the deployment. All these steps are shown in Figure 3.

1. *Classification of Metrics*

Following is the classification of metrics from the available metrics of the information technology infrastructure library (ITIL)

A. Time based

Availability: Total down time per service
Delay - Downtime divided by Uptime
Response times per incident
Actual availability compared with SLA requirements

B. Size based

Reliability – loss or error – Number of successful invocations divided by total

C. Combined (both time and size based)

Bandwidth – Tasks per time unit and average data blocks per time unit
Throughput – number of operations per second

D. Estimated attributes (historical or prediction)
CPU load, network load, free RAM, Free Disk space etc

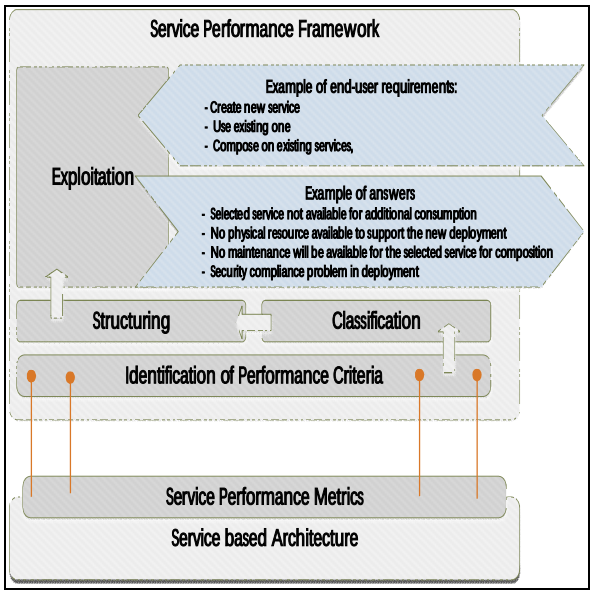


Figure 3. Proposed Approach

II. Structure of Metrics
Structure of metrics and key performance indicators has been designed with the help of ontology as shown in Figure 4.

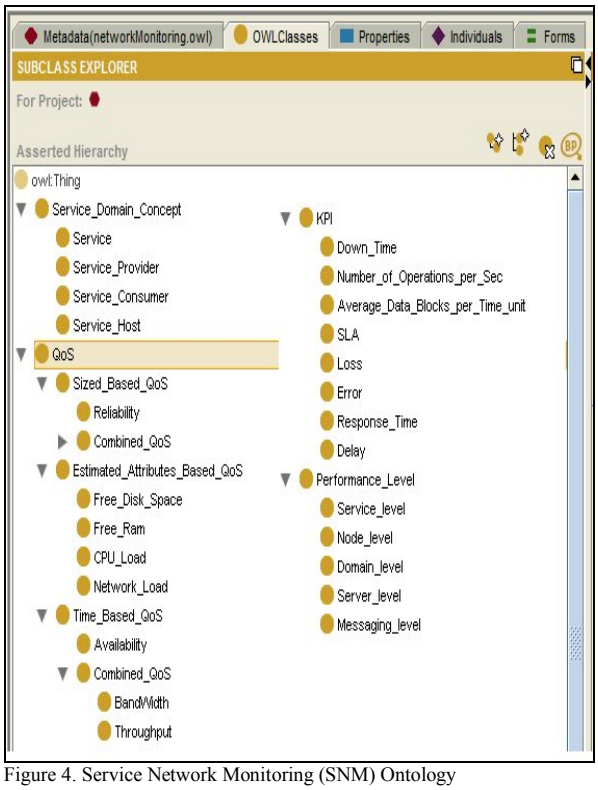


Figure 4. Service Network Monitoring (SNM) Ontology

A.

B. Service Network Monitoring (SNM) Ontology:
Service domain concept, QoS (quality of service), performance levels and KPI have is a relationship with the system. Then we have defined that service, service provider, service consumer and service host have is a relationship with service domain concept. Likewise time based QoS, size based QoS, combined QoS and estimated attributes have is a relationship with QoS as explained step I. Domain, node, service, service messaging and server have is a relationship with performance levels that have already been explained in section III. Similarly response time, delay, error, loss, SLA, number of operations per second and average data blocks per time unit have is a relationship with KPI(key performance indicators) as explained in the beginning of classification of metrics.

Domain	Node	Server	Service	Messaging
Loan organization	Customer accounting	Loan Organization Server	Check credit	Service messaging
Response Time	Loss	Network load	Response time	Response time per incident
Total Number of Users	Error	CPU load	Delay	Binding type
5 services	Delay	Requests Completed	Actual availability compared to SLA	No of bytes transferred
2 nodes	Response time	Average Request	No of operations per sec	No of operations per sec
Delay		Processing Time (seconds)	Loss	Average data blocks per time unit
		Available memory	Error	Loss / error

Figure 5. Key Performance Indicators

III. Decision Model
Decision model is explained with the help of pseudocode as shown in Figure 6. It takes ontology and user query as inputs and outputs will be selected service is not available, physical resource is not available for new deployment, maintenance of selected service is not available for composition and security compliance problem. First of all service has been selected that matches to the user query. If service is available then a function named as check KPI is called. If service is not available then a function named as check services for composition has been called. In this function all the services have been checked for service composition based on user query. If service composition is possible then a function named as check KPI is called. If service composition is not possible then a function named as create new service is called. In check KPI function all the KPI have been checked as shown in Figure 4. For each KPI, a comparison has been performed with SLA (service level agreements). If it is matched with the service level agreements then use that service as it is. If it is not matched then a new function named as check service status has been called. In check service status function service status has been checked. If service status

is success then three functions have been called in order to reach to the cause of the problem. The three functions are check protocol, check server and check nodes. In check protocol function details of the protocol that have been used to bind the service needs to be checked. If it is the source of problem then we need to change the messaging protocol. In check server, all the available resources that are provided at the server level have been checked. If they are the source of problem then we need to change the resources as shown in the estimated attribute node of the ontology. In check node function, service composition has been checked. If it is the source of problem then we need to change the service composition or create a new service.

```
Begin
        Input: User query, Ontology
        Output:
        Selected service ≠ Available
        Physical resource ≠ Available
        Maintenance of selected service ≠ Available for
        composition
        Security compliance problem
        Step 1:
        Select service that matches user query.
        Step 2:
        If Service = Available
        Then
                Check KPI
        Else
                Check for services for composition
        Step 3:
        Check for services for composition
        For each service = Available for composition
        Then
                Check KPI
        Else
                Create new service
        Step 3:
        Check KPI
        Step 4:
        For each KPI
                Compare from SLA
        If it is matched
        Then
                Use service or service composition as it is
        Else
                Check service status
        Step 5:
        Check service status
        If service status is success
        Then
                Check protocol
                Check server
                Check nodes
        Else
                Repeat step 2 for checking availability of other
                services
End
```

Figure 6: Pseudocode of Proposed Approach

CONCLUSION AND FUTURE WORK

In this paper we have proposed a framework to monitor the performance of service oriented systems to ensure their sustainability. First of all we have classified the performance metrics from the available metrics of the information technology infrastructure library (ITIL). Then we have designed our system ontology to show the relationships of all the concepts that we have used in our decision model. Finally a decision model has been designed in the form of pseudocode. Our next step will be to validate first case using "Oracle® Content Services Administrator"

REFERENCES

[1] K. Gottschalk, S. Graham, H. Kreger, and J. Snell. "Introduction to web services architecture," pp. 170-177, 2002

[2] Web Services Architecture, W3C Working Draft, 11 February 2004 – World Wide Web Consortium, http://www.w3.org/TR/ws-arch/

[3] Simple Object Access Protocol (SOAP) 1.2, Part 2, Adjuncts: (2007) – World Wide Web Consortium, http://www.w3.org/TR/soap12-part0/

[4] Web Services Description Language (WSDL) 2.0, part 1: Core Language (2007) – World Wide Web Consortium, http://www.w3c.org/TR/wsdl20/

[5] C. Peltz, "Web Service Orchestration and Choreography", Computer, IEEE, Issue No.10, vol.36, pp 46-52, October 2003

[6] Z. Kobti and W. Zhiyang." An Adaptive Approach for QoS-Aware Web Service Composition Using Cultural Algorithms", Advances in Artificial Intelligence Lecture Notes in Computer Science, Volume 4830, pp 140-149, Springer 2007

[7] F. Zulkernine and P. Martin. "Conceptual Framework for a Comprehensive Service Management Middleware", Advanced Information Networking and Applications Workshops (AINAW). 21st International Conference, pp 995-1000, May 2007

[8] G. Canfora, M.D. Penta, R. Esposito, M.L. Vilanni. "An Approach for QoS-aware Service Composition based on Genetic Algorithms," In: Proceedings of the 2005 conference on Genetic and evolutionary computation, pp. 1069–1075. ACM Press, New York 2006

[9] C. Zhou, L. T. Chia and B. S. Lee, "QoS-Aware Web Services Discovery with Federated Support for UDDI," Modern Technologies in Web Services Research, IGI Publishing Hershey New York.

[10] Z. Tari, A. K. A. Phan, M. Jayasinghe, V. G. Abhaya. "Benchmarking Soap Binding. On the Performance of Web Services," pp 35-58, Springer 2011

[11] Z. Tari, A. K. A. Phan, M. Jayasinghe, V. G. Abhaya. "The Use of Similarity & Multicast Protocols to Improve performance," On the Performance of Web Services, Springer pp 59-104, 2011

[12] Z. Tari, A. K. A. Phan, M. Jayasinghe, V. G. Abhaya. "Network Traffic Optimisation. On the Performance of Web Services," pp 105-138, Springer 2011

[13] E. Bertino, B. Carminati and E. Ferrari. "Authentication Techniques for UDDI Registries," Modern Technologies in web services research. IBM T.J. Watson Research, pp 9-30, USA 2007

[14] K. Curran and B. Gallagher. "Dynamically Adaptable Web Services Based on the Simple Object Access Protocol," Modern Technologies in web services research. IBM T.J. Watson Research, pp 54-75, USA 2007

[15] C. Mateos, A. Zunino and M. Campo." Mobile Agents Meet Web Services," Modern Technologies in web services research. IBM T.J. Watson Research, pp 98-121, USA 2007

[16] P. Fremantle, A Reference Architecture for the Internet of Things. Technical White Paper, version 0.8.0, 25 May 2014, http://wso2.com

[17] D. Lewis. "A Review of Approaches to Developing Service Management Systems", Journal of Network and System Management, pp 141-156, 2000

[18] Open Message Queue Developer's Guide for JMX Clients, Release 5.0. May 2013

[19] M. Richards, R.M-Haefel, and D.A. Chappell, Java Message Service second edition, ISBN: 978-0-596-52204-9, May 2009

[20] Systems Management: Application Response Measurement (ARM) API. Technical Standard C807, The Open Group, July 1998

[21] P. Bhoj, S. Singhal and S. Chutani. "SLA Management in Federated Environments" In proceedings of the sixth IFIP/IEEE International Symposium on Integrated Network Management, Boston, MA, 1999

Process performance measurement system for financial statements audit process in BPMS environment

Kristina Mijić*
* University of Novi Sad/Faculty of Economics Subotica/Subotica, Serbia
mijick@ef.uns.ac.rs

***Abstract*—Financial statements audit process is a key business process of each audit firm. In order to improve the business performances, audit firm should develop an adequate system of audit process performance measurement. In the first part, this paper defines performance measures for audit financial statements process, indicate to the problems and possible solution in order to create an adequate performance measurement system for audit process. The assumption of development an adequate system for measuring performance of audit process is the application of business process management concept into audit firm and development business process management software (BPMS) for the audit financial statement process. Furthermore, this paper presents the architecture of performance measurement system for audit process based on ODBC, MS Excel and BPMS. This kind of performance measurement system of audit process provides adequate and reliable information to managers for decisions making.**

I. INTRODUCTION

The aim of financial statements audit is to provide an opinion whether the financial statements are stated in accordance with financial statement regulation. Financial statements audit is a very important service which is provided by audit firms. According to the characteristics of financial statements audit such as a large number of activities, segregation of duties among the members of audit firm, defined document system etc. it can be concluded that the financial statements audit meets the definition and criteria of the business process. Business process is defined as a set of activities which use input and generate output creating value to the customer [5]. In the financial statements audit, auditors use the information about the audit client in order to obtain the evidence which will be use to create an audit opinion as the output of the audit process. The audit opinion in the form of audit report creates value for business decisions makers through increasing the reliability of financial statements as based for decision making.

Audit financial statement represents the crucial process for every audit firm, because more than half of revenue audit firm is realized by providing this service. The main objectives of the establishment of audit firms, in terms of the owners, are long-term business and profit. Competition in the audit market creates a special attention for audit firms to measure the performance of individual audit process. Performance is defined as an accomplishment of a given task measured against presently knows standards of accuracy, completeness, cost and speed [11].

The main objective of measuring process performance is gathering comprehensive and timely information on the performance of audit process. This information can be used to communicate goals and current performance of a business process directly to the process team, to improve resource allocation and process output regarding quantity and quality, to give early warning signals, to make a diagnosis of the weaknesses of a business process, to make a decision whether corrective action are needed and to assess the impact of action taken [12].

This paper focuses on performance measurement systems of financial statements audit process. Furthermore, this paper presents a set of performance measures of audit financial statements, the problems of development process performance measurement system for the purpose of audit process and possible solution for creation an adequate performance measurement system for audit process in BPMS environment.

II. PERFORMANCE MEASURES OF AUDIT FINANCIAL STATEMENTS - THEORETICAL APPROACH

Performance measures are the vital signs of the organization which quantify how well the organization achieves a specified goal [8]. When we are talking about specific types of performance indicators for business process there are no unified set of performance measures, because there is so many differences between business process among different organization. Every organization has to define the performance measurement system as a set of metrics which will be used to quantify efficiency and effectiveness of business process. Performance measurement system for business process can be characterized as an information system which:

- gathers information about the business process through a set of relevant performance indicators
- compares the current values against historical or target values
- provides the results (current value, historical or target value, gap, trend etc.) to managers for purpose of decision making in order to improve business process performances [13].

Adequate process performance measurement system shall involve the key indicators which will measure and describe the quality, time, flexibility and cost of business process [9]. For the audit firm it is necessary to define a

set of performance indicators at the level of audit process. Based on the results of performance measurement of audit process, the management of audit firms can take action for business process improvement in order to achieve a specified goal or even better business results. There are no unified adopted indicators to measure performance of audit process. The best indicators of performance audit process are those that indicate whether and how the implementation of the audit process contributes to the achievement of the objectives of the audit firm and whether it creates value for the audit firm.

International audit organizations propose various performance indicators of audit process in their frameworks and recommendations. According to the Audit Commission in London, performance indicators of audit process should include:

- measures of compliance of audit process with regulatory framework,
- the time required for the implementation of the audit process, the relationship between the planned and actual time spent on the implementation of the audit engagement,
- quality of audit staff (professional degrees and training) [1].

According to the Audit Committee Leadership Network in North America, audit process performance measures include:

- the costs of conducting the audit engagement,
- measures of assessing whether the audit procedures carried out in accordance with the plan of audit activities and regulatory framework,
- the time required for the implementation of the audit process,
- revenue per audit engagement,
- quality of audit process, which can be measured based on the time of engagement of every group of auditors (assistant, auditors and certified auditor) which are involved in the audit process [2].

III. SOLUTION OF PROCESS MANAGEMENT SYSTEM FOR AUDIT PROCESS

A. Problems and possible solution for development adequate and reliable process management system for audit process

Nowadays, realization of audit process in an efficient way is not possible without the use of IT/IS support. Auditors in order to conduct their activities are using various software, such as generalized audit tools, that are used to automate audit test, as well as software for document processing, communications etc. The use of different software solutions that are not integrated with each other, can lead to distortions of data integrity and integrity of audit process. In such circumstances it is not possible to adequately set and measure performance indicators for audit process, because the different software solutions are not integrated and do not follow audit process from start to end activity. For example how to ensure adequately measuring the time of the audit activities when the auditor works in the office of the audit firm as well in the office of audit client?

In order to create reliable system of performance measurement for audit process, implementation of BPMS in audit process will provide an adequate solution. BPMS can be defined as "a (suite of) software application(s) that enables the modeling, execution, technical and operational monitoring, and user representation of business processes and rules, based on integration of both existing and new information systems functionality that is orchestrated and integrated via services"[6]. The implementation of BPMS in audit process is possible because financial statements audit is business process according to the definition and criteria of business process [3]. The BPMS provide a wide range of benefits such as increasing the productivity and quality of business process. Furthermore, BPMS provide appropriate organization and storage of data about process through data warehousing. In order to gathered information about business process to become useful, it is important to create adequate system for measurement business process performance and reporting according to process performance indicators. BPMS is often applied technology for business process management, primarily because of their flexibility. The flexibility which BPMS offer can be noticed in several aspects. These systems provide a very simple way of creating business process model, which is later used to execute particular instances. On that occasion there is no need for any coding, so even the person without any programming experience is able to define and change the model [4].

B. The architecture of ODBC

The data about the realization of business process using BPMS are stored in appropriate database (MySQL, PostgreSQL, Microsoft SQL Server, Oracle etc.). The raw data about the conducted audit process should be transformed to information about the business process performance according with the defined set of performance indicators. The information of conducted audit process, or performance indicators of audit process, should be distributed to the management for the further analysis. The system for measuring and reporting on the performance indicators of audit process in the BPMS environment can be developed using the ODBC and MS Excel application.

ODBC (Open Database Connectivity) is defined as a standard programming language middleware API which is use for accessing database management systems (DBMS). ODBC provides a readymade layer between database and client side application (for example: MS Excel) [14]. How to connect the database of BPMS and client side application via ODBC? In order for ODBC to work, an operating system-specific Driver Manager needs to be utilized. A Driver Manager dynamically determines which ODBC driver to use for a program to access a database that is ODBC–compliant. The ODBC driver takes the request from the calling program, translates it to a native format that the database can understand, and the database performs the request [5]. ODBC architecture has four components:

- Application. Performs processing and calls ODBC functions to submit SQL statements and retrieve results.
- Driver Manager. Loads and unloads drivers on behalf of an application. Processes ODBC function calls or passes them to a driver.

- Driver. Processes ODBC function calls, submits SQL requests to a specific data source, and returns results to the application.
- Data source. Consists of the data the user wants to access and its associated operating system, DBMS, and network platform (if any) used to access the DBMS [10].

Fig. 1 shows the architecture of ODBC.

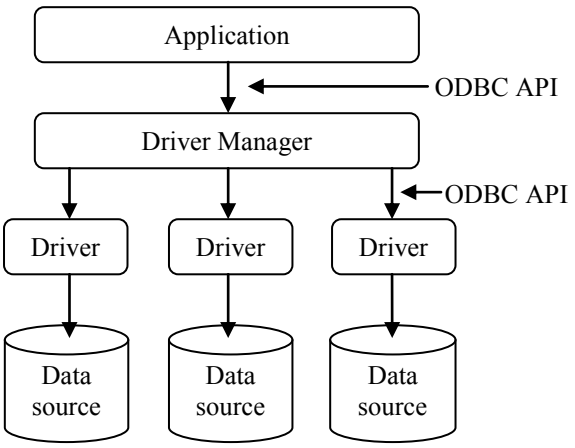


Figure 1. The architecture of ODBC [7]

For the purpose of audit firm "X" from Serbia, business process management software for audit financial statement was developed with open source solution ProcessMaker. Process performance measurement system was developed using ODBC and MS Excel application according to previous defined architecture. BPMS ProcessMaker are using MySQL database, so mysql connector odbc was used to make connection between database and MS Excel workbook (PMS Financial Statement Audit.xls). A simplified representation of architecture for performance measurement system for audit process is showed on following figure.

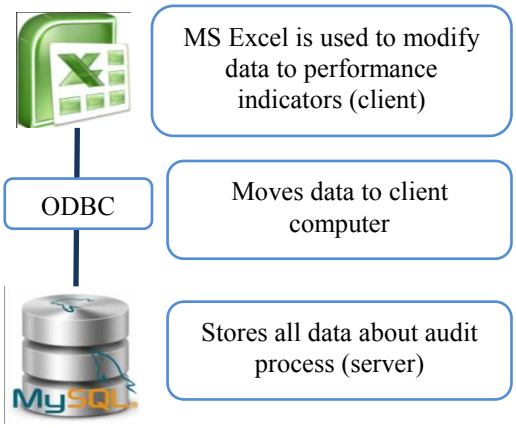


Figure 2. The architecture of data transfer from database to process management system created in MS Excel.

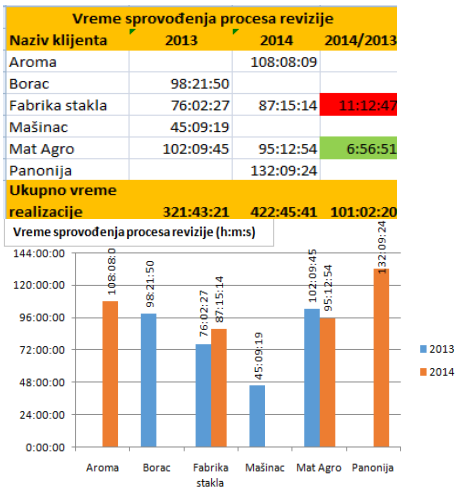
C. Performance indicators (measures) of audit process - practical review

Developed database of BPMS for audit process collects about 2,000 data of each audit process (on average). These data refer to general information about audit process, audit client, contract status, audit fee (income), the type of audit activity, time for realization every audit activity, data about audit stuff etc. According to these data, the following group of performance indicators was set:

- performance indicators of audit process efficiency,
- performance indicators of audit process quality,
- performance indicators of audit team quality,
- performance indicators of revenue from audit process.

Performance indicators of audit process are presented in tables in absolute or relative unites (depends on the type of indicator). Furthermore, performance indicators indicate to trends and gap (as a different between current value and target or average value). Also, in order to ensure simple and reasonable reading and analyzing, performance indicators are presented in graphs.

The group of performance indicators of audit process efficiency involves following indicators: the time of realization of audit process and audit activities, the time of work of each auditors. Fig. 3 shows report about performance indicators of time of realization audit process.



Vreme sprovođenja procesa revizije			
Naziv klijenta	2013	2014	2014/2013
Aroma		108:08:09	
Borac	98:21:50		
Fabrika stakla	76:02:27	87:15:14	11:12:47
Mašinac	45:09:19		
Mat Agro	102:09:45	95:12:54	6:56:51
Panonija		132:09:24	
Ukupno vreme realizacije	321:43:21	422:45:41	101:02:20

Figure 3. Performance indicators of audit process - time of realization of audit process

The time of realization of audit process can be present for one year or during the selected period. If managers of audit firm select presentation of time of audit process during the years, these performance indicators will also present the gap between two years. If the time for realization of audit process in 2014 (for example) is for 10% higher than in previous year, a red flag will be appear. On the other hand, if the time for realization of audit process is lower than in previous year, a green flag will be appear.

For each audit activity and audit task which is realized, the time of realization according to user (auditors) can be measured. If some auditors spent more than 110% of average time for realization some task, a red flag will be

shown. If some auditors spent less than 90% of average time, than the green flag will be appear (see fig. 4)

Vreme realizacije aktivnosti po izvrsiocima		
Izvršilac	A04 - Popunjavanje upitnika o prihvatanju klijenta	Odstupanje od proseka
Krpic Jovana	0:22:16	0:10:25
Lakovic Ljubinka	0:36:19	0:03:38
Malesevic Dusan	0:38:03	0:05:22
Radovic Marko	0:34:05	0:01:24
Prosečno vreme realizacije aktivnosti	0:32:41	

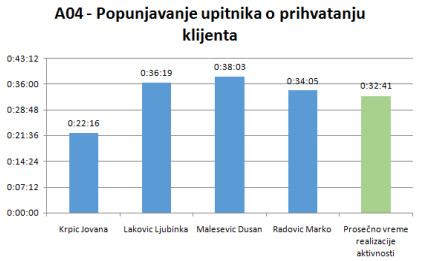


Figure 4. Performance indicators of audit process - the time of realization of specific audit task according to users (auditors)

The group of performance indicators of audit process quality include: the number of missing documentation per audit process, number and type of audit task which is realized but it auditors do not have an obligation to do. Next figure shows the number of missing audit documentation. If there is only one missing document, a red flag will be appear, because this is indicator that audit process was not implement totally according to audit regulation.

Nedostajuća dokumentacija 2013 godina			
Naziv klijenta	Izjava o nezavisnosti	Ugovor o reviziji	Ukupno
Borac	0	1	1
Fabrika stakla	0	0	0
Mašinac	0	0	0
Mat Agro	1	1	2
Ukupno	1	2	3

Figure 5. Performance indicators of audit process - the number of missing documents

The quality of audit teams can be measured according to structure of audit team, time of engagement of each single auditor, and time of engagement of group of auditors. The information about the structure of audit team and the time of engagement of auditor in audit process are presented on the fig. 6 and fig. 7.

Struktura revizijskog tima 2013 godina				
	Rukovodi			
Naziv klijenta	Ovlascen i revizor	lac revizije	Asistent revizije	Ukupno
Borac	1	1	1	3
Fabrika stakl	1	1	3	5
Mašinac	1	1	2	4
Mat Agro	1	1	2	4

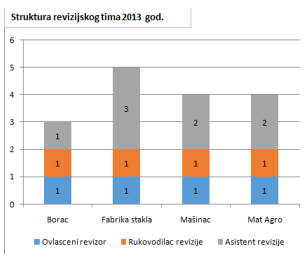


Figure 6. Performance indicators of audit process - the structure of audit team

Struktura revizijskog tima 2013 godina				
Naziv klijenta	Ovlascen i revizor	Rukovodilac revizije	Asistent revizije	Ukupno
Borac	32:09:08	25:09:12	41:03:30	98:21:50
Fabrika stakl	17:01:09	12:23:27	46:37:51	76:02:27
Mašinac	22:09:12	19:09:26	3:50:41	45:09:19
Mat Agro	26:06:27	22:43:18	53:20:00	102:09:45

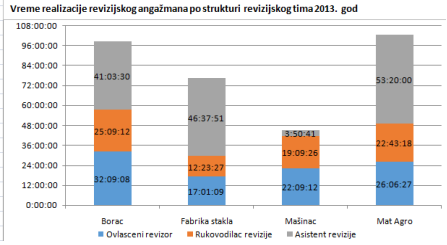


Figure 7. Performance indicators of audit process - the time of engagement of auditor in audit process

According to data about the audit fee, management of audit firms can easily measures contribution of single audit process to total revenue (see fig. 8)

Prihod po revizijskom angažmanu u dinarima za 2013 godina		
Naziv klijenta	Prihod	% Ukupnog prihoda
Borac	390,000	19.90%
Fabrika stakla	470,000	23.98%
Mašinac	400,000	20.41%
Mat Agro	700,000	35.71%
Ukupan prihod	1,960,000	100.00%
Prosečan prihod	490,000	

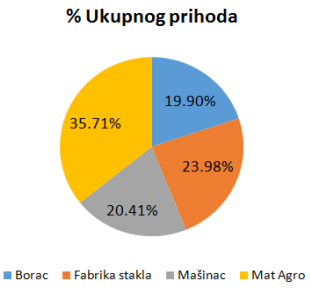


Figure 8. Performance indicators of audit process - audit fee

Comparative analyze between audit processes and some target value, as trend analyze of realization of audit processes, shall provide the timely and reliable information to managers of audit firms in function of decision making and business performance improvement.

These are some crucial performance indicators for audit process developed for audit firm in Serbia. Beside of these groups of performance indicators, which were defined by management of audit firm, also some other indicators were developed based on general information about audit processes. For example for managers of audit firm it is important to track the status of audit processes, and this can be done with audit processes dashboard (see fig. 9).

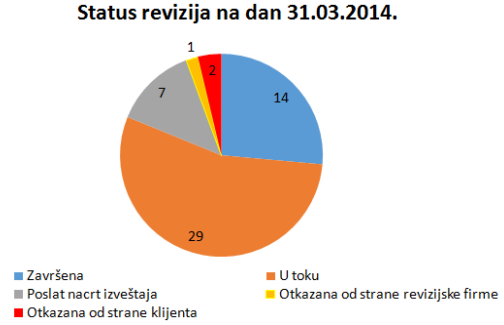


Figure 9. Performance indicators of audit process - the status of audit processes

CONCLUSION

Financial statement audit is a key business process of each audit firm, because the revenue of audit process presents more than half of total revenue. Regard to the specific characteristics of audit process, development of BPMS for audit process create opportunity for development of adequate performance measurement system. The BPMS for audit process management is developed by using open source BPMS solution. In order to create reliable, simple and low cost performance measurement system for audit process in BPMS environment, it was used ODBC and MS Excel application. This paper present architecture of performance measurement system for audit process developed by using ODBC and MS Excel, according to previous defined performance indicators. This kind of performance measurement system involves more than 20 audit process performance indicators, as well indicators according to recommendation of professional audit organizations. Also this system is very easily to be used and provide all necessary information about audit process to management for analyze the contribution of audit process in achievement organization goals and decision making.

REFERENCES

[1] Audit Commission, *Assessing the Effectivenes of External Audit.* AC: London, 2011.

[2] Audit Committee Leadership Network in North America, *Evaluating the Audit and the External Auditor*. ACLN: USA, 2011.

[3] D. Jaksić, K. Mijić, "The determination of selection factors of BPMS for the financial statements audit process," *Proceedings of III International Symposium Engineering Management and Competitiveness,* Zrenjanin, pp. 329-333, 2013.

[4] D. Mišić, M. Stojković, N. Vitković et al. "The concept of the information system for managing business processes of designing and manufacturing of oseofixation material," *ICIST 2014 - International conference on information society and technology - Proceedings,* pp. 10-16.

[5] J. Chang, *Business Process Management Systems,* Auerbach Publication, NY, 2006.

[6] J. Ravesteyn and J. Versendaal, Success Factors of Business Process Management Systems, *ACIS Proceedings of the 18th Australisian Conference on Information Systems.* Toowoomba, Austraila, 2007.

[7] K. Seifedine, "An implementation of ODBC web service," *Journal of Theoretical and Applied Information Technology,* vol. 26, No. 1, April 2011.

[8] K. Seokjin and N. Behnam, "The dynamics of quality costs in continuous improvement, " *International Journal of Quality and Reliability Management.* Vol. 25, No. 8, pp.- 842-859, 2009.

[9] Lj. Milanović Glavan, "Understanding Process Performance Measurement System, *"Business Systems Research,* vol. 2, No 2, pp. 25-38, 2011.

[10] Microsoft developer network, *ODBC Architecture,* Retrieved September 20, 2014 from http://msdn.microsoft.com/en-us/library/ms710238%28v=vs.85%29.aspx

[11] P. Bierbusse and T. Siesfeld. "Measures that matter," *Journal of Strategic Performance Measurement,* Vol. 1, No. 2, pp. 6-11, 1997.

[12] P. Kueng, "Supporting BPR through a Process Performance Measurement System, " *BITWorld Information Technology Management, Conference Proceedings.* pp.- 422-434, 1998.

[13] P. Kueng, "Process performance measurement system: a tool to support process based organization, " *Total Quality Management,* Vol. 11, No. 1, pp. 67-85, 2000.

[14] R. Khurana, *Connecting live data,* Retrieved September 15, 2014 from http://exceldashboards360.com

An Approach to Business Improvement by the Development of an Information System

Zoran Nešić*, Nebojša Denić **, Jasmina Vesić Vasović*, Miroslav Radojičić*
* University of Kragujevac, Faculty of Technical Sciences, Čačak, Serbia
** Faculty of Information Technology, Belgrade, Serbia
zoran.nesic@ftn.kg.ac.rs, denicnebojsa@gmail.com, jasmina.vesic@ftn.kg.ac.rs, miroslav.radojicic@ftn.kg.ac.rs

***Abstract* — This paper presents a methodological approach to the development of intelligent decision support systems in small and medium enterprises which respond to the requirements of modern management businesses via sophisticated information technology tools. The analysis has been performed with the purpose of introducing business intelligent systems into Serbian small and medium enterprises dealing with the sale of products and the provision of after-sales services. A segment of the system development has been presented on the specific example through the process of database creation.**

I. INTRODUCTION

It is increasingly necessary for managers of Serbian enterprises to be provided with information for conducting their activities of business decision-making in order to successfully and efficiently achieve organizational goals through the process of planning, organizing, leading and controlling the resources which are available for the organization [1]. It is obvious that small and medium-sized enterprises in Serbia will be faced with business changes in the business environment and such conditions that, in an increasingly competitive global business environment, often lead to changes and different needs for information for the business decision making process in companies [2]. Known authors have defined the functions of the high quality of the decision-making process in small and medium-sized enterprises in this direction. Mador (2002, p.4) primarily focuses on the importance of the rationality, scope and speed of decision making [3]. Filinov (2003, p.3) [4] indicates that solving the problem of decision making in a company depends on the type of management, the type and structure of the problem and the types of choices as to why certain decisions are made. The project development of intelligent business systems in small and medium-sized enterprises can be defined as the union of a sequence of complex and interrelated activities which have their goal or purpose. It must be carried out within a specified period, with a limited budget and must be in accordance with predetermined specifications [5]. The project is a temporary venture; therefore, in order to create a unique product, service or result, each project has its defined beginning and end. Wysocki & McGary, (2003 p.7-9) [5] argue that a project is determined by five parameters, i.e. the width, the quality, the price, the time and resources (human, financial, etc.). An investment is an important part and the right way to have a project implemented so as to introduce and implement business intelligence in small and medium-sized enterprises. In the context of the introduction of business intelligence, there are two different approaches that are commonly used. The first is the incremental or linear approach, in which approach each phase of the project is monitored. When the phase of the project is complete, the next phase is started. So, different phases of bringing a project to an end are monitored. The main drawback of this approach is its long duration, so users only receive the final solution, which means that errors and deficiencies are only discovered at the end, when the elimination of negative factors is complex and usually costly [6]. Such an approach is difficult. At the beginning of the project, future users of business intelligence should be well-defined. It is particularly suitable when an external contractor introduces an intelligent business system in accordance with specific requirements and specifications [7]. Many authors (Atre and Moss, 2003 [8]; Adelman and Mos, 2007 [9]; Howson, 2008 [7]) recommend the iterative approach to developing business intelligence in small and medium-sized enterprises as it is a much more flexible one (Sabherval & Becerra - Fernandez, 2010, p. 230-231) [6].

Small and medium enterprises opt for the most appropriate approach supported by a variety of factors, such as the size and complexity of the system, the resource availability and the like. A decision to apply the most appropriate approach is encouraged by various factors, such as the size and complexity of the system, resource availability etc. Apart from that, smaller companies usually have small budgets allocated for information technologies. In their practice of introducing business intelligence, small and medium-sized enterprises have often chosen a middle path, combining the features of both approaches [6]. The introduction and implementation of intelligent business systems include development tools and techniques for collecting, storing and accessing data for decision making at various levels of the company. Such sophisticated solutions are often costly, complex and long-term projects that require significant volume sources [10]. In this context, traditional approaches focus on the technical aspects and elements, while newer ones put an emphasis on the business impact [11].

The reason for the introduction of business intelligence into small and medium-sized start-ups is the business value, where such an introduction will be beneficial for the company. In the economic context, the commercial value of the investment is expressed as the net present value of cash flow after the deduction of tax per individual investment [11]. Also, Williams and Williams (2006, p. 12) [11] believe that we should take an attitude in spite of numerous business benefits business intelligence brings, that not every type of the business value of investment in

business intelligence can be assigned, while an advantage does not result in an increase in cash flow after tax deduction. Business intelligence is an area where traditional business-value-assessing techniques, particularly financial criteria, are not conducted well since many business advantages business intelligence brings are of a strategic type, for which very reason they are also difficult to measure [12]. There is a high probability that business benefits at a strategic level are the most important effects of the acquisition business intelligence brings [13].

II. DEVELOPMENT OF AN INFORMATION SYSTEM FOR SALES AND AFTER-SALES MONITORING

Further in this paper, the development of decision-making supporting information systems in small businesses in Serbia is presented, with the task to show the original and innovative process of creating the database of the company "Simonida"-Gračanica. The company specializes in selling and distributing automobile equipment and parts for large passenger cars and commercial vehicles on the territory of Kosovo and Metohija. As a tool for creating the database, the relational database management system for MS Access 2007 and Visual Studio 2008 has been used, being appropriat for the information system of small and medium-sized enterprises such as the company "Simonida". The presented approach is the development of a prototype system in a small company that does not have financial resources to initiate the introduction of expensive systems. It has been shown that this approach is introduced the fastest and that it immediately produces satisfactory results that their own intellectual resources can be utilized for the purpose of upgrading and maintaining itself, without additional financial costs. The management of "Simonida" Gračanica have conducted a detailed analysis and concluded that more information is necessary for business decisions given the increased workload. Auxiliary software solutions for decision support systems are massively available within the Microsoft Office package (MS Access and MS Excel) as well as sophisticated and complex tools such as Visual Studio 2008, which are particularly suitable as a solution to a large number of small and medium-sized enterprises in Serbia, which significantly reduces operating costs, enables the successful operations of the company and generates a profit. Considering that the economic and business development in Serbia is based on small and medium-sized enterprises, subject matter of the analysis carried out in this paper represents opportunities for the development and implementation of new software solutions to decision support in these companies. The business world implies the way in which people communicate, represent, transmit and share their knowledge with others in order to improve their operations and the achievement of common goals. Managers and employees can go through information about the market, customers, competitors, partners' internal activities, products and services in order to contribute to the creation of the business value and improve the business impact [14].

"Simonida's" sales of cars and spare parts are a challenging and complex activity. Among the cars, there are many differences such, as whether the car has a gasoline engine or used diesel fuel, whether it has power steering etc. Figure 2 shows the algorithm of the business processes in the sales of new vehicles.

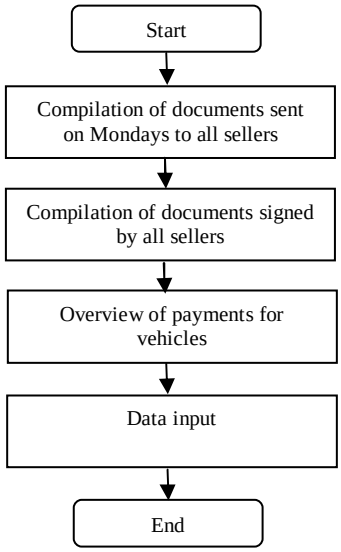


Figure 1. Monitoring the sale of new cars [15]

The biggest problem is the buyer's desire r to buy a car with a certain type of equipment. In order for "Simonida's" sellersor sellers on auto markets to know which car has a certain type of equipment, they should know what each car separately has. The customer would spend a lot of time to find a car that suits him. Because of this, it was necessary to find a practical solution to searching "Simonida" company's cars.

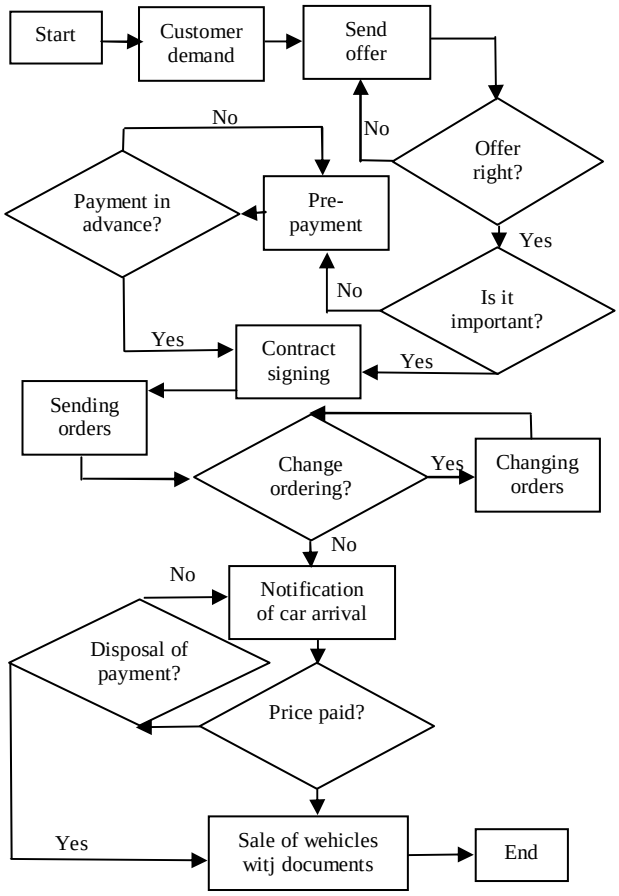


Figure 2. The algorithm of the business processes [15]

The solution lies in creating a database of the company that will contain all the information about the cars and equipment they have. It is necessary to make an interface, i.e. the type of the software program which will enable customers to quickly locate the desired car. The program should be able to perform the searching of the entire database of the company "Simonida" and to present the results to the user. The program should be as simple as possible so that the customers who have never used anything of that kind could use very quickly.

III. DATABASE

Before the very beginning of the development of decision support systems in "Simonida"-Gračanica, a company for car sales and the distribution of spare parts and equipment, a detailed analysis of the company was made. the company's needs were taken into account and the best ways to have those needs realized were discussed. There are three tables that have been created in the databas: the ***Car Table*** – which contains information about cars. The ***Customer Table*** – which contains information about customers. The ***Transactions Table*** – which contains information about completed transactions.

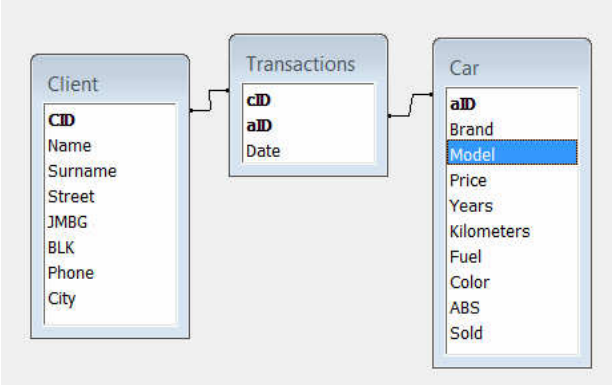


Figure 3. The relational database model [15]

The ***Car Table*** consists of the following fields: ***aID*** – the primary key table and the unique identifier for each car in the database; ***Brand*** – the car brand, ***Model*** – the model of car, ***Price*** – the car price, ***Years*** – the year of the car production, ***Kilometers*** – the car mileage, ***Fuel*** – the type of the car fuel, ***Color*** – the car color, ***ABS*** - additional equipment, ***Sold*** – whether the car has been sold or not.

The ***Client Table*** consists of the following fields: ***Name*** **–** the customer personal name, ***Surname*** – the customer surname, ***Street*** – the address of the buyer, ***JMBG*** – the personal unique identification number of the customer, ***BLK*** – the ID card number of the customer, ***Phone*** – the phone number of the customer, ***City*** - the city where the buyer lives.

The ***Transactions Table*** consists of the following fields: ***aID*** – the unique identifier for each car in the database. ***Date*** - the time when the transaction was completed.

The information solution implemented by using the "Microsoft Visual Studio 2008" program allows the user to perform the basic interaction with the database. These are the following options: input, edit, search and delete cars and clients. Additionally, it enables the storage of the client's transactions.

IV. APPLICATION SOLUTION

The application consists of several forms, the most important ones being as follows:: ***Form1.cs*** – the form gives the user an opportunity to review all cars in the database and open all other forms of the application. This functionality is implemented by using datagridview control. ***FrmNew.cs*** – the form gives the user the option of adding a new car in the database. ***frmUpdate.cs*** – the form gives the user an option to change the car data. ***FrmSearch.cs*** – the form gives the user the ability to search the car database. ***FrmList.cs*** – the form that displays a list of clients, where the user has an option to delete data or obtain the forms for editing a particular client. ***FrmCustomers.cs*** – the form gives the user the option of adding a new client and a new transaction in the database. ***FrmClient.cs*** – the form gives the user an option to change data in the client base. ***Transaction .cs*** – the form provides the user with the ability to preview the list of all transactions, print those lists and delete individual transactions.

The datagrid is the main control on the Form1.cs. Each row represents one car. One row is composed of 9 columns: ***ID*** – the field that displays the number of the car. ***Brand*** – the field that displays the brand of the car. ***Model*** – the field that displays the model of the car. ***Price*** – the field that displays the price of the car. ***Sold*** – the field that displays the status of the car. ***Shopping*** – the button which, when pressed, opens the form FrmKupci.cs. ***Delete*** – the button which starts the process of deleting the selected cars from the database.

Figure 4 shows the design of the "Form1.cs" form. In the header of this form, control is used, consisting of the following elements: ***The new car*** – displays the frmnew.cs form. ***Search*** – displays the frmSearch.cs form. "Transactions" shows the "Transaction.cs" form. "Clients" – shows the "FrmList.cs" form, and "Close program" – closes the program.

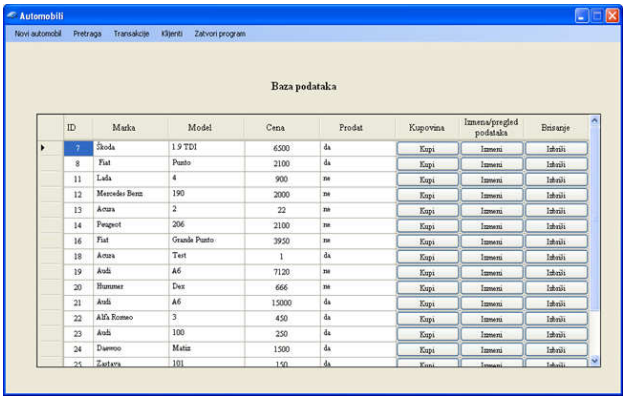


Figure 4. The form for the review of all cars [15]

The "frmUpdate.cs" form (Figure 5) allows changes of the basic characteristics of the cars and includes the following controls: "Mark" – the list of the predefined cars. The user is only obliged to choose a particular brand of the car when entering data changes for the car. "Model" – the standard text field where the user enters the car model. "Price" – the standard text field where the user enters the price of the car. "Year" - a list of predefined years. "Mileage" – the standard text field where the user enters the mileage of the car. "Fuel" – the list of the

predefined fuel. "Color" - the standard text field where the user enters the color of the car. "Sold"– the standard text field where the user enters the status of the car. The control for changing accessories consists of 15 "checkbox" controls. The same validation is applied as the form for adding a new car, i.e. all the fields in the control for changing the basic characteristics of the car are mandatory whereas all the fields from the other sections are optional.

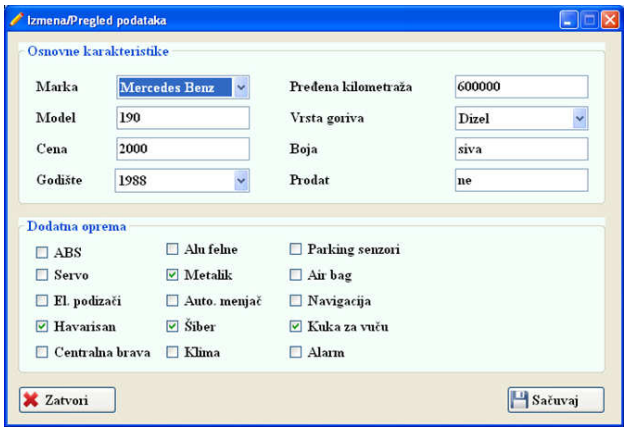


Figure 5. The form for editing and changing data [15]

Figure 6. The form for the registration of customers [15]

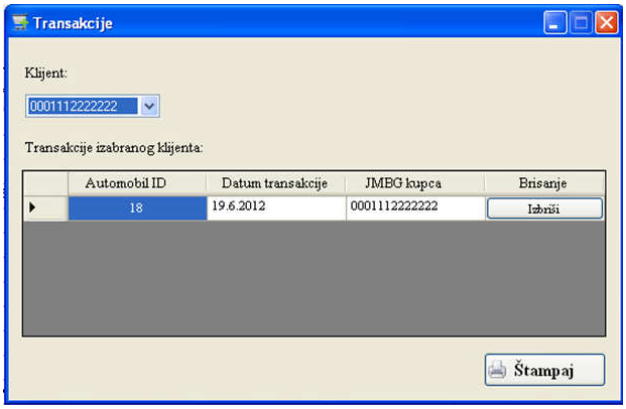


Figure 7. The form for transactions [15]

The "FrmCustomer.cs" form (Figure 6) is displayed by the "Buy" action on the Form1.cs form. The main control includes the following controls: ***Name*** – the standard text field where the user enters the first name of the buyer. ***Last Name*** – the standard text field where the user enters the last name of the customer. ***The City*** – the standard text field where the user enters the city of the customer. ***Street*** – the standard text field where the user enters the the customer's street. ***Identification number*** – the standard text field where the user enters the ID number of the customer. ***BLK*** – the standard text field where the user enters the ID card number of the buyer. ***Telephone*** – the standard text field where the user enters the customer's phone number.

The "Transaction.cs" form (Figure 7) is displayed when the user activates the "Transactions" link in the main navigation on the "Form1.cs" form. The Form consists of the following controls: "Client combobox" – so programmed that this list only shows the ID number of the clients that are in the database. When the user selects an ID number in the list in the "datagrid", the control displays the selected client's transactions. Each transaction consists of four columns: "Auto ID" – the unique identifier of the purchased cars; "Transaction Date" – the date of the completion of the transaction; "Customer JMBG" – the personal unique identification number of the buyer. "Delete" – deletes transactions from the "Transactions" table.

Figure 8. The form for the registration of clients [15]

The "FrmList.cs" form (Figure 8) is displayed by using the "Clients" link in the main navigation on the Form1.cs form. Each row in the control stands for a single client. One line consists of six columns: ***Name*** – the field that displays the first name of the client; ***Last Name*** - the field that displays the surname of the client; ***Street*** – the field that displays the client's street; ***JMBG*** – the field that displays the personal unique identification number of the client. ***Edit/view data*** – the button which opens the "FrmClient.cs" form. ***Delete*** – the button that starts the process of deleting the selected client from the database.

V. EFFECTS OF APPLICATION AND FURTHER IMPROVEMENT

After the improvement and implementation of the IS and after having performed the analysis according to the particular target group of customers, their requirements in purchasing vehicles, their expectations and the distribution and sale of spare parts, the "Simonida" company for the sale of vehicleshas adjusted its range of vehicles and spare parts. Based on the parameters, amongst which the specifications of the vehicles and spare parts are one of the most important ones, and according to the customer's interest and the time (certain periods of the year), the company has adapted itself to their business; with such a range of customized sales, significant savings in unnecessary procurement (storage) costs of vehicles and spare parts have been achieved. Applying the IS after the six-month financial report, the company for the sale of vehicles "Simonida" has realized income in the sales of vehicles and spare parts and has also reduced the

associated costs (of the transport of vehicles, the storage rental of space for parking vehicles etc.).

The graphic report in Figure 9 accounts for the growth of the vehicle sales and profits in relation to the period prior to the implementation of the company's information system.

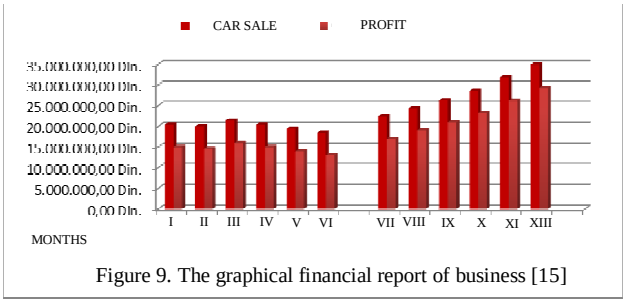


Figure 9. The graphical financial report of business [15]

From the above-presented graph, it can be concluded that the information system of the company "Simonida", a company dealing with the sale of vehicles, only contains parts for an entry, search and edits of the data about the vehicles, the customers and the transactions. The improvement of the information system can be accomplished by introducing the so-called modules for decision support i.e. by creating graphical representations and reports such as:

- The report on the sale of the vehicles in the selected time interval
- The report on the total sales of certain brands of vehicles
- The report on the customers and the like

The proposed dimensions of the OLAP (On-Line Analytical Processing) cubes for the implementation of the elements of business intelligence and creating a previous report are shown in Table 1.

Table 1.

The description of the OLAP cube dimensions [15]

Cube – CarSale	
Dimension	Description
Time	The dimension constructed on the basis of the ***date*** field from Fact_SellingVehicles based on the data source; the dimension only includes data on the days when the sale was made. The hierarchy of the dimensions can be determined in the year, quarter, month, and day.
Car	The Car Dimension allows the monitoring of vehicle sales. The hierarchy of the dimensions can be determined according to the brand and model of the car.
Client	The Clients Dimension is created on the basis of the "Name" field in the Customers table. The hierarchy of the dimensions can be determined by the city.

Also, taking into account the fact that the considered company is engaged in the distribution of spare parts, a module for recording spare parts and equipment can be added. On the basis of transactions, an analysis of the client's needs for certain parts and equipment can be carried out, on the one hand, and, on the other hand on the basis of the results of the analysis, a strategy related to the purchase of equipment and spare parts that clients usually claimed can be created.

VI. CONCLUSION

This paper presents the key segments in the development of the information system of the company "Simonida", a company dealing with the sale of vehicles and aftersales. The present economic results of the company after the introduction of the information system are indicative of the positive impact on its overall business. The practical results of the application of the present system have shown the improvement of the data management in the enterprise and that it can satisfy all the needs for information necessary for the management of the enterprise. In addition, this has provided a connection of the key business functions (finance, accounting, distribution, manufacturing and sales). The flow of information between those key business functions ensures that the enterprise has a complete and full control over the business, which provides it with effective management and decision making, which gives an additional aspect of applicability. Thanks to its features and functions, the decision-making process has been facilitated and accelerated, which increases the competitiveness and has a positive impact on the efficiency of operations and creates significant savings for this company.

The advantages of the above information system are, among other things, reflected in the fact that the processes of entry, editing, deleting and searching are very easy and enable a quick and easy use of this system. To add, it allows a great interaction with the end-user. This paper presents a concrete example that can universally be applied as a starting point in the formation of an information system with companies engaged in car sale and after-sales.

REFERENCES

[1] R. L. Daft, ***Management,*** South-Western College Pub, 2007

[2] P. Poon, and C. Wagner, "Critical Success Factors Revisited: Success and Failure Cases of Information Systems for Senior Executives," ***Decision Support Systems,*** Vol. 30, pp 393-418, 2001.

[3] M. Mador, "Strategic Decision Making Processes: Extending Theory to an English University", 2002 Available at: http://ecsocman.edu.ru/images/pubs/2002/12/25/0000033000/str_des_making.pdf

[4] N. B. Filinov, "Business Decision-Making in the Era of Intellectual Entrepreneurship", 2003, Available at: http://www.wspiz.pl/~unesco/articles/book3/tekst7.pdf

[5] S. Williams, and N. Williams, ***The Profit Impact of Business Intelligence***, Morgan Kaufmann Publishers, San Francisco, 2007.

[6] R. Sabherwal, and I. Becerra-Fernandez, ***Business Intelligence: Practices, Technologies, and Management***, Wiley, Hoboken: NY, USA, 2010.

[7] C. Howson, ***Successful Business Intelligence: Secrets to Making BI a Killer App.,*** McGraw-Hill Osborne Media, 2007.

[8] L. T. Moss, and S. Atre, ***Business Intelligence Roadmap: The Complete Project Life cycle for Decision-Support Applications,*** Addison-Wesley Professional, Boston: MA, USA, pp 77-81, 2003.

[9] S. Adelman, L. Moss, and M. Abai, ***Data Strategy,*** Addison Wesley: United States, 2007.

[10] R. K. Wysocki, and R. McGary, ***Effective project management: traditional, adaptive, extreme,*** Wiley Pub: Indianapolis, 2003.

[11] M. Valle, "Visualization and art", 2006. Available at: http://www. isedj. Org /isecon/2007 /2523/ISECON.2007.Segall.pdf http://www.cscs. ch/~mvalle/visualization/VizArt.html

[12] Z. Irani, and P. E. D.Love, "The Propagation of Technology Management Taxonomies forEvaluating Investments in Information Systems", ***Journal of Management Information Systems,*** Vol. 17, No. 3, pp 161-177, 2000.

[13] M. Gibson, D. Arnott, I. Jagielska, and A.Melbourne, "Evaluating the Intangible Benefits of Business Intelligence: Review & Research Agenda", Proceedings of the 2004 IFIP International Conference on Decision Support Systems (DSS2004): Decision Support in an Uncertain and Complex World, pp. 295-305, 2004.

[14] D. Marchand et al., Mastering Information Management Harlow, ***Financial Times***. Prentice-Hall, UK, pp. 295-300, 2000.

[15] Internal corporate documents of the company “Simonida” Gračanica

.

Scheme for mapping scientific research data from EPrints to CERIF format

Valentin Penca*, Siniša Nikolić *, Dragan Ivanović *

* University of Novi Sad/Faculty of Technical Sciences/Department of Computing and Automatics, Novi Sad, Serbia
{valentin_penca,sinisa_nikolic, chenejac}@uns.ac.rs

***Abstract*— This paper describes basics of the EPrints institutional repository and CRIS systems and their data models. The result of this research is mapping scheme of the data from EPrints to the CERIF standard.**

I. INTRODUCTION

Rapid development of science and technologies resulted with huge amount of various data. One of the most important tasks will be how to store and make data accessible. Institutional Repository (IR) can resolve the mentioned issue. In [1], an IR is described as an electronic system that captures, preserves and provides access to the digital work products of a community.

The three main objectives for having an institutional repository are:

- To create global visibility and open access for an institution's research output and scholarly materials.
- To collect and archive content in a "logically centralized" manner even for physically distributed repositories.
- To store and preserve other institutional digital assets, including unpublished or otherwise easily lost ("grey") literature (for example, theses or technical reports).

The availability of open-source technologies affect on the rapid development of IRs worldwide, particularly among academic and research institutions. Therefore, it is not surprising the existence of several open source software platforms available for developing IRs like *EPrints* [2], *DSpace* [3], *Greenstone* [4], *Fedora* [5] and *Invenio* [6]. *IR EPrints* was the first open-source repository software to be developed. In [7] is stated that commonly used institutional repository systems are DSpace and Eprints. Paper [8] advocates that EPrints is a powerful and inexpensive solution for sharing scholarly works with the world. Drawback of all mentioned IR is that they have their specific metadata models causing difficulties in data exchange between diverse systems. One possible solution is using the Common European Research Information Format (CERIF) standard [9], which is the basis of Current Research Information Systems (CRISs), for exchanging data from scientific-research domain. In this paper the scheme for mapping data about research from EPrints IR to CERIF format is proposed. That scheme can be used as a guideline, supporting the exchange between *EPrints* repositories and CRIS systems. Motivation for this work was also to extend and improve research from [10].

II. EPRINTS IR

EPrints is an open-source software package for building open access Institutional repositories. Software was developed at the University of Southampton School of Electronics and Computer Science and released under a GPL license. This presents clear advantages for institutions with smaller budgets and that have programmers on staff. EPrints was the first IR software packages to appear and has been available for 14 years. EPrints is a Web and also command-line application based on the LAMP [11] architecture (but is written in Perl rather than PHP). It successfully runs under multiple OS platforms, like Linux, Solaris, Mac OS X and Microsoft Windows. Version 3 of the software introduced a (Perl-based) plug-in architecture for importing and exporting data. Current stable release is 3.3.11/31 from January 2013. EPrints is fairly interoperable [12], supporting OAI-PMH [13] and SWORD [14].Comparison and advantages of EPrints to other IR repositories is described in [15]. According to the official data presented in [16] there are over 535 EPrints registered instances running all over the world which are part of ROARMAP (Registry of Open Access Repositories Mandatory Archiving Policies).

Word "eprints" in particular means “electronic publications” (papers, lectures, videos, etc). Therefore, EPrints is a piece of software for managing "eprints". Basic entity in EPrints is the *Data Object (DataObj)*, which is a record containing metadata [17] and has unique identifier. In EPrints exists three core objects (Figure 1) *EPrint*, *Document*, *User*. All core objects extend *DataObj*. Between Core Objects are defined some relations like: one *User* owns (deposit) many *EPrints*, or one *EPrint* has many documents attached to it. In EPrints one or more documents (files) can be linked with the data object (*Document*). File objects are an interface to file-system files (or cloud buckets) which can be downloaded to allow a person to read a publication (eprint).

A *DataObj* is a collection of "metadata fields". There are many different types of metadata fields. Type affects how a field is rendered, indexed, searched and so forth. Every field has a *type*, *name* property and indicator that states if the field is multiple or not. *Data Object*s are usually stored in the database. They are generally spread over a number of tables containing with the same prefix. Every *Data Object* has system fields (which are set by the system, and not alterable), but the *User* object and *EPrint* object have additional fields which are configured on a per-repository basis. These non system fields can be customized in the Perl script files user_fields.pl and

eprint_fields.pl. Detailed information of Metadata Field Types and their configuration can be found in [18].

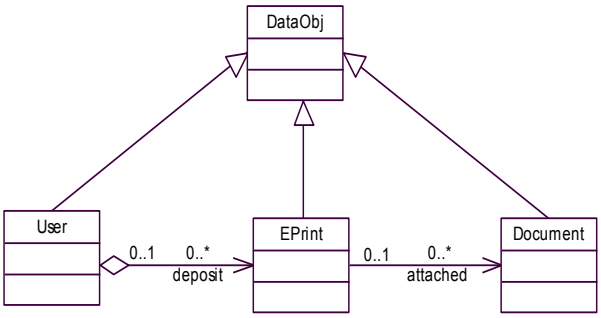


Figure 1 - EPrints data model

We've described the individual *Data Object* but a repository holds many eprints, documents and has many registered users. For that reason a concept of *dataset* for collection of *DataObj* is included. A dataset represents all data objects of a certain type in a single repository. All data objects in the repository are part of the three core datasets "eprint" (all eprints), "user" (all registered users) datasets, "document" (all document). Multiple fields (lists of values) are stored in their own table named after the dataset, then an underscore followed by the fieldname, e.g. "eprint_subjects". These tables also contain a "pos" value to indicate the order of the list.

III. CERIF MODEL

CERIF is the standard which describes data model that can be used as a basis for an exchange of data from scientific-research domain. CERIF Standard describes the physical data model [19] and the exchange of XML messages between the CRIS systems [20]. The best feature of CERIF is that it can be expanded and adapted to different needs. In practice, CERIF is often mapped to other standards that also represent the data of scientific-research domain, for example CERIF/MARC21 mapping described in [21]. Authors of [22] recommend an extension of CERIF that incorporates a set of metadata required for storing theses and dissertations. Another example is [23] where authors argue how CERIF can be used as a basis for storage of bibliometric indicators.

Hereinafter we will present main entities of the CERIF data model version 1.5

- Base Entities - represent the core (basic) model entities. There are only three basic entities cfPerson, cfOrganizationUnit and cfProject.
- Result entities - A group of entities which includes results from scientific research like publications, products and patents. Representatives of this group are: cfResultPublication, cfResultProduct and cfResultPatent.
- Infrastructure Entities - represent a set of infrastructure entities that are relevant for scientific research. The entities which belong in this group are: cfFacility, cfEquipment and cfService.
- 2nd Level Entities - Entities that further describe the Base Entities and Result Entities. E.g. cfEvent is one of those entities, stating the event.
- Link Entities - are used to link entities from different groups. Typical entities of this group are: cfOrganizationUnit_OrganizationUnit, cfOrganizationUnit_ResultPublication and cfResultPublication_DublinCore. Link Entities allow for a generic classification mechanism to define their meaning, indicating the role for each entity instance in a relationship. Every Link entity is described with a role (*cfClass*, *cfClassScheme*), timeframe of relation (*cfStartDate*, *cfEndDate*), value(*cfFraction*) and identifiers of elements creating relation (e.g. *cfOrgUnit*, *cfResPublId*). The 'role' in link entities is not stored directly as attribute value, but as reference to Semantic layer.
- Multiple Language Entities - These entities provide multilingualism in CERIF for some entities.
- Semantic Layer Entities - Provide different kinds of semantics in CERIF model. The entities in this group are cfClassificationScheme and cfClassification. Those entities are used to describe classes and classification schemes for link and other entities. CERIF prescribes a controlled vocabulary to describe some of the classifications.
- Additional Entities - Currently in this group are classified entities that represent DC record.

Figure 2 [Figure 2] shows some of Base, Result, Link and Multiple Language Entities which are relevant for the mapping proposed in this paper.

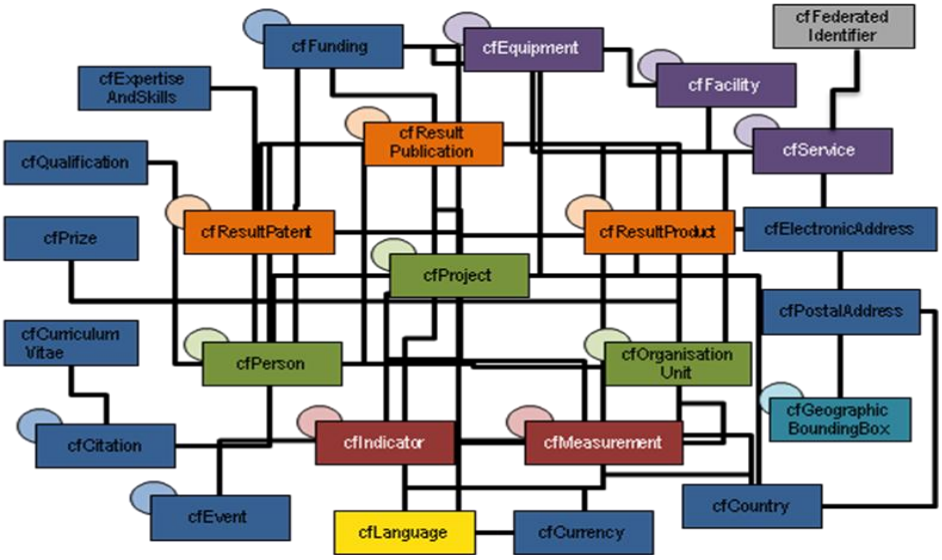


Figure 2 – CERIF model

IV. EPRINTS WITH OTHER SYSTEMS

Storage of data from another system in EPrints was performed in [24]. This paper presents mapping of data from Artext system whose primary users are a national and international community of artists, art historians and curators. In order to obtain the data from Artext, it was necessary to customize default EPrints distribution, adding the new fields and/or modify existing ones. Mappings to EPrints data model was carried out also in [25]. The ReCollect plug-in for EPrints was released at the end of the Research Data @Essex project in 2013. Based on the customization work undertaken while developing a pilot research data repository at the University of Essex, ReCollect offers a low barrier route to deploy such a service, allowing a deposit, review, publication and presentation of research data collections.

The necessity of mapping data between EPrints and systems based on CERIF model is completely natural when we know that these two systems are keeping similar data from scientific research domain [26]. EPrints IR has a large number of users in the UK since it was created at the University of Southampton. In the UK, universities and other research organizations are obligated to prepare themselves for the next national research assessment exercise, known as REF [27]. For that reason a REF2014 EPrints plug-in has been specifically designed to help institutions. REF plug-in [28] is special XML based application profile called CERIF4REF. The mentioned XML is foreseen as a document that is based on CERIF model. EPrints repositories have a potentially important role to play in institutions' responses to the requirements of REF [29].

EPrints contains the plug-ins for the export and import in various formats such as CSV, HTML, DC, ATOM, JSON, MODS, METS. One of the existing plug-ins allows exporting the data in CERIF 2008 1.0 XML format. That CERIF plug-in allows exporting only of some kind of data like books and articles publications to CERIF XML format.

V. MAPPING SCHEME FOR EPRINTS TO CERIF

The motivation of authors for mapping EPrints data to CERIF data model is found in the fact that they are part of the development team of CRIS UNS system [30], which currently does not have the ability to acquire data from the EPrints system. In the paper [31], a CERIF compatible research management system CRIS UNS is presented, which can be accessed at http://www.cris.uns.ac.rs. Currently, the system stores over 10,500 records of scientific publications (papers published in journals, conference proceedings, monographs, technical solutions and patents etc.). CRIS UNS system is under development since 2008 at the University of Novi Sad in the Republic of Serbia. Former development of that system covered implementation of the system for entering metadata about scientific research results [32]. Later phases in the development of CRIS UNS system included integration of various extensions that relay on CERIF model.

Also, the motivation was to include all the *Eprint* data object entities that are not covered with existing CERIF Export plug-in, in accordance with the newer CERIF 1.5 version. Proposed mapping scheme did not include any customization of EPrints distribution since the most users use just the default installation.

Every *EPrints* object has a property called *type* that determines the *EPrints* object type in system (eg. article, book, exhibition, video, etc). The value of that property sets which metadata fields are actually in use from *that* object. So, it represents one of the essential information for mapping. Table 1 defines the mapping to CERIF entities according to the types of *EPrints* object. For easier understanding, all *EPrints* types of objects are grouped and classified into general types of scientific research data (Scientific research data type). For every general type there is an adequate CERIF entity on which it can be mapped. Initial step of mapping is done by resolving the type of *EPrints* object and creating the instance of appropriate CERIF entity.

TABLE I.
EPRINTS OBJECTS CLASSIFICATION

Scientific researh data type	EPrints type	CERIF entity
Publications types	article book_section monograph conference_item book thesis	cfResPubl
Patent types	patent	cfResPat
Event types	exhibition performance	cfEvent
Product types	artefact composition image video audio dataset experiment	cfResProd
Other	teaching_resource other	Non Mapped

All the scientific research data in EPrints is stored as the values of metadata fields. Proposed mapping of all available metadata fields to CERIF format is available at [33]. Due to space limitations, only a segment of mapping is presented in Table 2 and Table 3. For the purpose of explaining the mapping concept, several EPrints metadata fields are selected. Table 2 defines a list of EPrints fields for which the mapping is possible. Also, within that table the EPrints fields' properties are explained. In addition to the field name (*EPrints metadata field*), there is a column *options* where a list of field possible values is noted. For example, field *event_type* states the supported classification of an event as: *conference*, *workshop* and *other*. Fields where the *options* column is empty may have an arbitrary value. Column *multiple* indicates that the field can be found several times in EPrints record (eg. record has more than one author). Column *EPrints type* defines which types of *EPrints* Objects (Table 1, column 2) use the corresponding field (e.g. EPrints field *series* is used only by EPrints type *book*). For some fields that can be applied to more than one *EPrints type* a general group name (Table 1, column 1) or keyword *all* is used. Thus, for the *creator* filed (used in context of for any kind of *EPrints* object) the value of the column *EPrints type* is *all*. The last field in the Table 2 *CERIF entity* defines for which of CERIF entities (single entity or a list) listed in Table 1 the appropriate EPrints field can be utilized. Value *All* indicates that the EPrints field can be used for any CERIF entity that is identified in Table 1 column 3, e.g. each of the specified entities have a *creator*. Value *NONE* is used when it is not possible to perform the mapping for EPrints metadata field to adequate CERIF entity (e.g. *learning_level*).

CERIF model relies on *Link* and *Semantic Layer Entities* to provide additional semantic between entities and for some particular entities. So, it is to be assumed that a large portion of metadata fields from EPrints object will be stored as instances of those CERIF entities. Table 3 presents CERIF entities and their attributes (enclosed in brackets) that are used in mapping. There are three distinct cases of mapping. First, in which the EPrints metadata field is directly stored as an attribute of CERIF entity (e.g. metadata field *series*). The value of that field is stored within attribute *cfSeries* of *cfResPubl* entity. Second case is the situation where the EPrints metadata field can have one of possible values defined in column options. For every possible value, an appropriate CERIF classification is necessary. Every created CERIF entity instance and its classification are connected with the identifier (e.g. *cfResPublId*, *cfPersId* etc). In this scenario column *CERIF core, result and 2nd level entities* specifies instance of CERIF entities for which link entities are defined in column *CERIF link entities. Classification* for those link entities are stated in column *Used CERIF classification*. For example, EPrints metadata field *ispublished* with value *inpress* is classified with CERIF scheme *cfResPubl_Class* and class *In Press*. Entity *cfResPubl* and *cfResPubl_Class* are connected with identifier *cfResPublId*. The most complex is the third scenario, where one value of EPrints metadata field is usually mapped to more than one CERIF entity. This scenario requires the creation of entities from columns *CERIF core, result and 2nd level entities* and *CERIF link entities.* Also, the adequate classification for *CERIF link entities* is defined in column *Used CERIF classification*. Third scenario will be explained with metadata field *creator*. At first, for *creator* an instance of core entity *cfPers* needs to be created. The value of *creator* field will be stored within attributes *cfFamilyNames* and *cfFirstNames* of entity *cfPersName*. *cfPersName* is

TABLE II.
EPRINTS METADATA FIELDS PROPERTIES

Eprints metadata field	options	multiple	EPrints type	CERIF enity
creators		X	all	All
event_type	conference workshop other		Conference item	cfEvent
subjects		X	all	cfResPubl cfResPat cfResProd
learning_level			Teaching resource	NONE
ispublished	pub inpress submitted unpub		publications types	cfResPubl

TABLE III.
EPRINTS METADATA FIELDS MAPPING

Eprints metadata field	options	CERIF core, result and 2nd level entities	CERIF link entities	Used CERIF classification
creators		cfPers (cfPersId) cfPersName_Pers (cfPersId,cfPersNameId) cfPersName (cfPersNameId,cfFamilyNames, cfFirstNames)	Publications types: cfPers_ResPubl (cfPersId,cfResPublId) Patent types: cfPers_Res (cfPersId,cfResPatId) Event types: cfPers_Event (cfPersId,cfEventId) Product types: cfPers_ResProd (cfPersId,cfRedProdId)	scheme:cfPers_ResPubl, class:Author scheme:cfPers_ResPat, class:Patentee scheme:cfPers_Event, class:Performer scheme:cfPers_ResProd, class:Constructor
ispublished	pub inpress submitted unpub	cfResPubl(cfResPublId)	Publications types:cfResPubl_Class (cfResPublId)	scheme:cfResPubl_Class, class:Published scheme:cfResPubl_Class, class:In Press scheme:cfResPubl_Class, class:Submitted for Consideration scheme:cfResPubl_Class, class:Unpublished
series		cfResPubl(cfSeries)		

connected to *cfPers* by link entity *cfPersName_Pers*. The *creator* field is integral part of information about publications, patents, events and products. Thus in CERIF, *cfPers* can be linked with *cfResPubl*, *cfResPat*, *cfEvent* and *cfResProd*. In case when *creator* is an author of publication, linking is done within the entity *cfPers_ResPubl*. For stating the role "author of publication", a CERIF semantic scheme *cfPers_ResPubl* and class *Author* are utilized.

VI. CONCLUSION

The importance of institutional repositories and CRIS systems for scientific research data is enormous. Making data accessible between these systems is unavoidable. Therefore, this paper presents mapping scheme for EPrints data to CERIF model for CRIS systems.

The main contribution of this research is:

• Proposal for mapping data from EPrints repository to the current 1.5 CERIF model

• Potential possibility for creation of new or expansion of existing CERIF-XML Export plug-in

Future work will be directed towards mapping the data from other IRs like Fedora, Greenstone and Invenio to CERIF format.

ACKNOWLEDGMENT

Results presented in this paper are part of the research conducted within the Grant No. III-47003, Ministry of Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] N. F. Foster and S. Gibbons, “Understanding faculty to improve content recruitment for institutional repositories,” -Lib Mag., vol. 11, no. 1, pp. 1–12, 2005.

[2] “EPrints - Digital Repository Software.” [Online]. Available: http://www.eprints.org/. [Accessed: 20-Dec-2014].

[3] “DSpace | DSpace is a turnkey institutional repository application.” [Online]. Available: http://www.dspace.org/. [Accessed: 20-Dec-2014].

[4] “Welcome :: Greenstone Digital Library Software.” [Online]. Available: http://www.greenstone.org/. [Accessed: 20-Dec-2014].

[5] “Fedora Repository | Fedora is a general-purpose, open-source digital object repository system.” [Online]. Available: http://fedora-commons.org/. [Accessed: 20-Dec-2014].

[6] “Invenio.” [Online]. Available: http://invenio-software.org/. [Accessed: 20-Dec-2014].

[7] J. Kim, “Finding documents in a digital institutional repository: DSpace and Eprints,” Proc. Am. Soc. Inf. Sci. Technol., vol. 42, no. 1, 2005.

[8] E. Sponsler and E. F. Van de Velde, “Eprints.org Software: a Review,” Jul. 2001.

[9] “Common European Research Information Format | CERIF,” 2000. [Online]. Available: http://www.eurocris.org/. [Accessed: 18-Jan-2014].

[10] V. Penca and S. Nikolić, “Scheme for mapping Published Research Results from Dspace to Cerif Format,” in 2. International Conference on Information Society Technology and Management, 2012, pp. 170–175.

[11] “ONLamp.com.” [Online]. Available: http://www.onlamp.com/. [Accessed: 20-Dec-2014].

[12] M. Castagné, "Institutional repository software comparison: DSpace, EPrints, Digital Commons, Islandora and Hydra," Library, Archival and Information Studies (SLAIS), School of, Aug. 2013.

[13] "Open Archives Initiative Protocol for Metadata Harvesting." [Online]. Available: http://www.openarchives.org/pmh/. [Accessed: 20-Dec-2014].

[14] "SWORD." [Online]. Available: http://swordapp.org/. [Accessed: 20-Dec-2014].

[15] J.-G. Bankier, Institutional Repository Software Comparison. UNESCO, 2014.

[16] "Registry of Open Access Repositories." [Online]. Available: http://roar.eprints.org/view/software/eprints.html. [Accessed: 20-Dec-2014].

[17] L. YOGESH and P. NEELIMA, "OPEN SOURCE DIGITAL LIBRARY SOFTWARE (OSS-DL): ASSESSMENT AND EVALUATION," Int. J. Libr. Sci. Res. IJLSR, vol. 3, no. 3, pp. 21–30.

[18] "EPrints Metadata Fields Documentation." [Online]. Available: http://wiki.eprints.org/w/Category:EPrints_Metadata_Fields. [Accessed: 20-Dec-2014].

[19] B. Jörg, K. Jeffery, J. Dvorak, N. Houssos, A. Asserson, G. van Grootel, R. Gartner, M. Cox, H. Rasmussen, T. Vestdam, L. Strijbosch, V. Brasse, D. Zendulkova, T. Höllrigl, L. Valkovic, A. Engfer, M. Jägerhorn, M. Mahey, N. Brennan, M. A. Sicilia, I. Ruiz-Rube, D. Baker, K. Evans, A. Price, and M. Zielinski, CERIF 1.3 Full Data Model (FDM) Introduction and Specification. 2012.

[20] J. Dvořák and B. Jörg, "CERIF 1.5 XML - Data Exchange Format Specification," 2013, p. 16.

[21] D. Ivanović, D. Surla, and Z. Konjović, "CERIF compatible data model based on MARC 21 format," Electron. Libr., vol. 29, pp. 52–70, 2011.

[22] L. Ivanovic, D. Ivanovic, and D. Surla, "A data model of theses and dissertations compatible with CERIF, Dublin Core and EDT-MS," Online Inf. Rev., vol. 36, no. 4, pp. 548–567, 2012.

[23] S. Nikolić, V. Penca, and D. Ivanović, "STORING OF BIBLIOMETRIC INDICATORS IN CERIF DATA MODEL," Kopaonik mountain resort, Republic of Serbia, 2013.

[24] T. Neugebauer, C. MacDonald, and F. Tayler, "Artexte metadata conversion to EPrints: adaptation of digital repository software to visual and media arts documentation," Int. J. Digit. Libr., vol. 11, no. 4, pp. 263–277, Dec. 2010.

[25] T. ENSOM, "RECOLLECT: TECHNICAL OUTPUTS FROM THE RESEARCH DATA @ESSEX PROJECT," presented at the JISC MANAGING RESEARCH DATA PROGRAMME WORKSHOP, BIRMINGHAM, 2013.

[26] L. Carr, "EPrints: A Hybrid CRIS/Repository?," in CERIF and Institutional Repositories, Rome, 2010.

[27] "REF 2014." [Online]. Available: http://www.ref.ac.uk/. [Accessed: 20-Dec-2014].

[28] "EPrints - Are you ready for REF 2014?" [Online]. Available: http://www.eprints.org/ref2014/. [Accessed: 20-Dec-2014].

[29] A. Clements and V. McCutcheon, "Research Data Meets Research Information Management: Two Case Studies Using (a) Pure CERIF-CRIS and (b) EPrints Repository Platform with CERIF Extensions," Procedia Comput. Sci., vol. 33, pp. 199–206, 2014.

[30] "Current Research Information System of University of Novi Sad." [Online]. Available: http://www.cris.uns.ac.rs/. [Accessed: 18-Jan-2014].

[31] D. Surla, D. Ivanovic, and Z. Konjovic, "Development of the software system CRIS UNS," in Proceedings of the 11th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, 2013, pp. 111–116.

[32] D. Ivanović, G. Milosavljević, B. Milosavljević, and D. Surla, "A CERIF-compatible research management system based on the MARC 21 format," Program Electron. Libr. Inf. Syst., vol. 44, no. 3, pp. 229–251, 2010.

[33] "Mapping EPrints to CRIF." [Online]. Available: http://s000.tinyupload.com/?file_id=35285195853574771781. [Accessed: 29-Dec-2014].

Information Security Awareness through a Virtual World: An end-user requirements analysis

Christos Mettouris*, Vicky Maratou**, Divna Vuckovic***, George A. Papadopoulos*, Michalis Xenos**
*University of Cyprus, Nicosia, Cyprus
**Hellenic Open University, Patra, Greece
***Center for the Promotion of Science, Belgrade, Serbia
mettour@cs.ucy.ac.cy, v.maratou@eap.gr, dvuckovic@cpn.rs, george@cs.ucy.ac.cy, xenos@eap.gr

Abstract—**Living in the digital era, computers and the Internet became important tools used by people to support significant parts of their everyday life such as work, education, socializing, entertainment, communication, etc. However, there are certain risks involved in using ICT technologies, and thus all ICT users should be aware of the basic principles of information security and data protection. No matter how much expertise is put into securing information assets and networks (e.g. firewalls, encryption) the human factor always remains a vulnerability. Our vision is to aid towards the development of information security awareness culture by using a 3D Virtual World Learning Environment that will simulate real-life security threat scenarios, examples and counterexamples in a way that different groups of users will experience the risks and combine critical skills, knowledge and collaboration to overcome them, without exposing their organization to real risk. In this paper we provide the results of the end user requirements collection and analysis in order to define and develop the specifications of the aforementioned 3D Virtual World Learning Environment and the specifications of the in-world activities.**

I. INTRODUCTION

The V-ALERT LLP EU project [1] aims to support the establishment of an Information Security culture in different ICT user target groups (pupils and teachers, ICT students, academics and enterprise employees) by providing awareness through an innovative and immersive e-learning tool. An online 3D Virtual World Learning Environment (VWLE) will be developed which will simulate real-life Information Security threat scenarios, allowing users to gain first-hand experience of the different risks and threats, though in a safe manner. The 3D VWLE of V-ALERT will also provide real time in-world assistance to the users through personalised recommendations. The underlying pedagogy of the V-ALERT approach is the "learning by doing" which can increase intrinsic motivation of learners and lead to deeper understanding and learning [2, 3].

The aim of this work is to present the results of the end user requirements collection and analysis which will be used to define and develop the specifications of the 3D VWLE and the specifications of the in-world activities for the V-ALERT project. The requirements collection from the end users was accomplished through an online questionnaire which end users from 5 different European countries had to complete. Following the requirements collection, an analysis of the acquired data was conducted, the outcomes of which are presented in this study.

Section 2 of this paper discusses related work. In section 3 we describe the questionnaire's aim and content, while in section 4 we discuss the results obtained. Section 5 closes the paper with conclusions, as well as a list of important user requirements based on which the 3D VWLE, its activities and the security threat scenarios will be designed and implemented.

II. RELATED WORK

Compared to other e-learning technologies, 3D virtual worlds can provide learners with a full understanding of a situation using immersive 3D experiences which allow the learner to freely wander through the learning environment, explore it, obtain sense of purpose, act, make mistakes, collaborate and communicate with other learners [4]. Indeed, immersion, that is the feeling of "actually being there", accompanied with the interaction with virtual objects can enhance learners' interest and engagement to the learning tasks and help them to develop a stronger conceptual understanding, depending on the content [5, 6]. Therefore, with the prospect of providing learners with experiences they would otherwise not be able to experience in the physical world (or in a classroom), a rapidly growing interest in 3D virtual world learning activities is observed by a large number of schools and universities worldwide [3, 7].

A number of research works aim at providing learners with experiential learning of different scientific topics through simulations or role-play games in 3D interactive virtual worlds [3, 4, 8, 9]. However, there are not many works focused on delivering Information Security issues [10, 11, 12] through 3D virtual worlds. To this direction, mostly 2D simulations and games have been developed [13].

In the context of V-ALERT, we attempt to fully exploit the possibilities of the 3D virtual world technology to create and evaluate experiential learning simulations which will address the users' real needs for Information Security awareness. To this aim, through an organised requirements collection procedure, we acquired direct feedback from more than 600 of different, in age and expertise, end users, from 5 European countries. The results of the data analysis have revealed interesting issues for educators and 3D VWLE developers who are

interested in designing educational and learning sessions on Information Security.

III. End User Requirements Collection Via Questionnaire

A. Aim of the Questionnaire

The end users requirements collection was accomplished through an online questionnaire. To the survey participated users from Cyprus, Greece, Serbia, Croatia and Bulgaria. We have categorized our end users in 4 different target groups as follows: (i) students of primary or secondary education, (ii) teachers or academic professors, (iii) ICT college or university students and (iv) enterprise staff or employees in an organisation or administrative personnel.

The questionnaire included specific questions that aimed to acquire important information regarding the user profiles, interests and activities from the perspective of Information Security. Besides basic information such as gender, age, country, etc., the aim was also to collect useful information regarding computer/smartphone usage and social network and online user activities, as well as to understand the aspects of information security that are more important to the users, based also on the particular target group of each user. In this manner, we aimed to discover the needs, preferences, habits and the level of security awareness of each target group regarding information security.

Furthermore, the questionnaire aimed to lead the process of defining the conceptual specification of the 3D VWLE. More particular, the information obtained from the end users will point to the right direction regarding the appropriate learning approach and learning activity (e.g. role-playing gaming theory, etc.) to be adopted for each target group and in relation also to the different user "roles", activities and pedagogical objectives.

In order to have a statistically correct analysis, we had aimed for 100 responses per target group. Partners sharing the same target groups were to guarantee a sum of 100 participants together.

B. Questionnaire Content

The questionnaire was launched on the official website of the V-ALERT project [1]. 5 project partners were responsible for the end users and had 2 weeks to make sure that their end users had completed the questionnaire on time. The questionnaire included 42 questions, from which 40 were multiple choice questions; the two exceptions include stating age and country in text boxes. 5 questions included text boxes for the user to provide input other than the multiple choice answers.

The first 6 questions aimed to collect user basic info. Following, questions 7 to 10 were about how often users use computers/smartphones and how confident they are while using them. Questions 11 to 19 got more into detail about online shopping, e-banking and similar activities users may perform daily. Questions 20 to 27 attempted to acquire user information about activities users perform in their routine that may involve risks, such as exchanging sensitive information via the internet with others, as well as how secure users perceive their activities to be. Questions 28 to 35 were more technical and concerned their level of knowledge and education on security awareness. The last section of the questionnaire (questions 36 to 41) included questions about users' experience regarding 3D virtual worlds and about how they perceive the involvement of 3D virtual worlds in simulating real-life security threat scenarios for educational purposes. Finally, the last question (No. 42) and probably the most important one in the questionnaire, explicitly asked users about the type(s) of security threats they would like to learn more about, by offering a list of the 13 most well-known threats to select from, as well as a text box to add more.

IV. Questionnaire Data Analysis & Results

Our analysis is based on studying the results of all end users as one large dataset, as well as analysing the results per target group in order to identify the particular characteristics of each group. In this section we discuss the most important results of our survey, providing the results in a percentage approximation.

For simplicity, in the following we will refer to: "Students of primary or secondary education" as students, "Teachers and academic professors" as teachers/profs, "ICT College and University students" as univ. students and "Enterprise staff, employees in organisations and administrative personnel" as employees.

A total of 666 responses were acquired. 361 responses - 54.2% were students, 49 responses - 7.4% were teachers (12) and academic professors (37), 194 responses - 29.1% were univ. students and 62 responses - 9.3% were employees.

A. Computer Usage and Confidence

The first few questions were about how often users use computer-smartphones and how confident they are while using them. Most of them use a computer (PC, laptop) for online activities on a daily basis, as well as a mobile device such as a smartphone. Only a very small students' minority never uses a computer for online activities, while for mobile devices the corresponding percentage includes users from all 4 target groups. 1 out of 5 teachers/profs do not use mobile devices for online activities.

More than 75% of the users are confident in using a computer for online activities, while for mobile devices the percentage drops to 70%. From a small percentage that are not confident at all with computers most are students and employees. Regarding mobile devices, 1 out of 20 users from each target group do not feel confident at all.

Regarding online shopping, more than 40% of users declare that they do not shop online, but this is mostly because most of our participants are students. Only 3 out of 10 students shop online; regarding the other target groups, the percentage is 80-98%. Almost all employees (98%) seem to shop online. Most online shoppers shop from home. However, there are a few of the online shoppers that have the risky habit of shopping from anywhere; these are mostly employees and univ. students. This may happen due to added confidence these two target groups may feel. One possibility is that employees and univ. students trust the networks and their security

software (e.g. antivirus) without a real deep knowledge as to what threats they could really be exposed to.

Regarding e-banking and m-banking, 1/3 of the adult participants manage an account. Most people have e-banking or both. Very few participants have only an m-banking account. This can lead to the conclusion that, while adult users shop online and manage e-banking accounts, they do not really trust mobile devices to manage also an m-banking account. Or, maybe m-banking is not needed, since most of them already have an e-banking account. Again, employees seem to be more risky since those that use m-banking from any location they believe is secure are more (almost twice) than the employees that use m-banking from home. Moreover, the confidence rate regarding e-banking is relatively low in all target groups. Many people have stated that they are not confident at all while performing online transactions through e-banking, except from employees: none of them has stated that they do not feel confident.

B. Routine Activities

The next part of the questionnaire acquired user information about activities users perform in their routine that may involve risks, such as exchanging sensitive information via the internet with others, as well as how secure users perceive their activities to be.

Despite that, by now, every person using computational devices (computers and smartphones) must have repeatedly heard to never disseminate personal sensitive information over the internet (email, social networks), unfortunately our users do so. A few of them even do this on a regular basis with strangers, while more than 15% do it sometimes. When being asked whether they exchange sensitive information via the internet with family or friends, the percentages rise: almost 1 out of 10 users do it regularly and 1 out of 2 do it sometimes. It is important to note that employees do it at a much lower percentage than the others.

Moreover, there are many users (almost 1 out of 5) that do not believe that shopping online may risk the exposure of sensitive data from their side, and a further 1 out of 5 that do not know. These users certainly need to be informed and educated on online information security matters. It is important to note that, while most users belonging in the latter set of users are students and univ. students, nevertheless, this set includes users from the other 2 target groups as well.

Another important result is that 1 out of 2 users believe that their home internet connection is very secure while using an Ethernet cable, and more or less the same statistic applies also for a Wi-Fi wireless connection. Of course, in reality this is not the case.

Finally, users in general do not think that using a public computer is very safe, while 1 out of 3 have stated that they do not know, mostly students. Also, 1 out of 4 participants have stated that they know very well how to protect and secure their electronic data from cyber threats when using a public Wi-Fi. 1 out of 5 do not know at all how to do that and half of them are somewhere in the middle: don't know exactly how to do it but are aware of some protective measures.

C. Technical Quiz

The following 8 questions in the questionnaire (Q28 - Q35) were more technical, aiming to determine the participants' level of knowledge and education on security awareness. The mean percentage of correct answers for all participants regarding the technical questions was 49%. Only one person responded correct to all questions. Moreover, 44.8% of the participants responded correct to at least one question (mean value). The success rate of two questions that had multiple correct answers was very low (15% and 5%), although the encouraging thing here is that a large number of participants have answered partially correct: for one question 1 out of 3 participants answered 50% correctly and for the other one almost half of them answered 66% correctly.

On the rest of the questions, the participants were correct at a rate of 65-86%, except from Q29 ("While browsing the Internet you receive a message informing you that you have become a victim of spyware and that you should click 'OK' in order to remove it. What do you do?") and Q35 ("What is phishing?"). It seems that Q29 tricked most participants and hence the low success rate, while Q35 is low mostly because of the unsuccessful participation of the students. It is interesting to observe that, although Q29 is a relatively easy question, most participants stated that they would "close the browser and scan for spyware" instead of just "ignoring the message" (correct answer). Although their response is incorrect, it would not risk their safety. Also, it is important to note that in this question students were more successful than the other target groups, perhaps due to the simple and straightforward way children are able to think as opposed to adults.

One may assume that students would have lower success rates than other target groups regarding the technical questions. Students have a higher rate regarding incorrect answers in Q31 ("To your opinion, which of the following are characteristics of a strong password?") and a lower success rate than the other target groups in Q34 ("What is a computer virus?") and Q35, but they also have a higher success rate than the other target groups in Q29.

D. User Experience

The last section of the questionnaire determined the users' experience regarding 3D virtual worlds and about how they perceive the involvement of 3D virtual worlds in simulating real-life security threat scenarios for educational purposes. These questions were set on a 5-level Likert scale, where the participants were asked to state their opinion on how much they agree to the question/statement by selecting one of the following: "Strongly agree", "Agree", "Neutral", "Disagree" and "Strongly disagree". The option "Don't know" was also included.

Half of the participants have stated that they have previous experience with computer games. As expected, students and univ. students have the highest rate in computer games experience. 1 out of 5 participants do not have previous experience with computer games. Also, more than 40% of all participants have previous experience in 3D virtual worlds through a computer (PC,

mobile device). Students, univ students and employees have the most, while teachers/profs seem to have the lowest. Almost half of teachers/profs do not have previous experience in 3D virtual worlds through a computer. In this research, the opinion of participants that have previous experience in 3D virtual worlds is important, since, through their experience, these users have become the most relevant people to respond to our questions.

In the question whether users believe that 3D virtual worlds could be effectively used for educational purposes by offering educational oriented experiences to the user, almost 27% of all participants strongly agree, while 34% agree (therefore, a total of 61% agree). Moreover, from the participants that have stated to have previous experience in 3D virtual worlds through a computer, 76% believe that 3D virtual worlds could be effectively used for educational purposes by offering educational oriented experiences to the user, and only 6% do not. Also, 50% of users who do not have previous experience in 3D virtual worlds through a computer believe that 3D virtual worlds could be effectively used for educational purposes by offering educational oriented experiences to the user, while 19% do not. Another interesting point is that very few participants strongly disagree with the statement (3.6%), and a further 6.5% disagrees. 33% of these users have stated to have previous experience with computer games and 25% have stated to have previous experience in 3D virtual worlds through a computer. 13% have both experiences.

Regarding the different target groups and whether they believe that 3D virtual worlds could be effectively used for educational purposes, students agree by more than 50% (and disagree by 13%), univ. students agree by 67% (and disagree by 8%), teachers/profs agree by more than 70% (and disagree by 4%) and employees agree by 79%. None of the employees disagrees in any way with the statement, and furthermore, none of the teachers/profs strongly disagrees with the statement.

The next question was whether the participants believe that 3D virtual worlds facilitate a 'learning by doing' educational model or not (Q39). 52% of all users agree to the statement and less than 10% do not. From the users that have previous experience in 3D virtual worlds through a computer, 67% believe that 3D virtual worlds facilitate a "learning by doing" educational model and only 7% do not agree. 45% of users who do not have previous experience in 3D virtual worlds through a computer believe that 3D virtual worlds facilitate a "learning by doing" educational model, 18% do not. Students agree to the statement by 38% (and disagree by 12%), univ. students agree by 66% (and disagree by 7%), teachers/profs agree by 64% (and disagree by 4%) and employees agree by more than 40%. None of the employees strongly agrees to the statement, as well as none of them disagrees in any way with the statement. Also, none of the teachers/profs strongly disagrees with the statement.

Question 40 explicitly asked the users whether they would like to participate in learning sessions facilitated through 3D virtual world simulations. More than half of the users (59%) agree, 17% were neutral, 13% disagree and 8% did not know. From the users that disagree, 28% have stated to have previous experience with computer games, 24% have stated to have previous experience in 3D virtual worlds through a computer and 9% stated to have both experiences. From the users that have previous experience in 3D virtual worlds through a computer, 78% would like to participate in learning sessions facilitated through 3D virtual world simulations and only 7% would not like to participate in such learning sessions. 46% of the users who do not have previous experience in 3D virtual worlds through a computer would like to participate in learning sessions facilitated through 3D virtual world simulations while 28% would not. In each one of the four target groups more than half of the participants would like to participate in learning sessions facilitated through 3D virtual world simulations. Almost 57% of students would like to participate (and 16% would not), 61% of univ. students would like to participate (and 11% would not), 50% of teachers/profs would like to participate (and 12% would not) and 69% of employees would like to participate (and 5% would not). Moreover, 27% of users who do not have previous experience in 3D virtual worlds through a computer and believe that 3D virtual worlds could be effectively used for educational purposes by offering educational oriented experiences to the user, also believe that 3D virtual worlds facilitate a "learning by doing" educational model and furthermore would like to participate in learning sessions facilitated through 3D virtual world simulations.

Regarding Q41: *"If you have any previous experience in 3D virtual worlds through a computer (PC, laptop, mobile device), for what purpose was your participation in 3D virtual worlds: 1. Educational, 2. Gaming, 3. Recreational Sports & Rehabilitation, 4. Scientific visualization and 5. Other"?* The responses of all participants show that 41.4% was Gaming and 19.1% was Educational (the other purposes were less than 10%). The purpose of participation in 3D virtual worlds of the users that have stated to have previous experience in 3D virtual worlds was 56% Gaming and 19% Educational (the other were less than 10%). Regarding the different target groups, students', univ. students' and employees' participation was mainly gaming, while teachers/profs' participation was mainly Educational, as well as Gaming.

The final question of the questionnaire and a very important one as well, is Q42: *"What type of security threats from the list below would you like to learn more about and gain knowledge in order to avoid them?"* This was a multiple choice question with 15 possible answers, 13 of which included a threat (see bars in Fig. 1 and Fig. 2 – the y-axis depict percentages): 1. Identity Theft, 2. Cyber bullying, 3. On-line Sexual Harassment, 4. Social Networking Misuse , 5. Cyber Scams, 6. Phishing/Spam, 7. Unauthorized exposure of personal information to online social networks, 8. Misuse of internet access exposing devices security, 9. Breach of Intellectual Property Rights, 10. Information leakage, 11. Social Engineering Attacks, 12. On line web security threats and 13. Unauthorized physical access to corporate facilities

The results were the following (Fig. 1): most of the participants (62%) have selected Identity Theft (bar 1 in Fig. 1) as the type of security threat they would like to learn more about. Therefore Identity Theft is the most

important security threat for the users. Second in the list of user preferences is Cyber Scams (bar 5 in Fig. 1) with 50% of users selecting it, while third is Phishing/Spam (bar 6) with a similar rate (49.8%). Note that the second and third selections are well below the first one by an important gap (12%). Then we have Unauthorized exposure of personal information to online social networks (43.7%, bar 7), Information leakage (44.6%, bar 10), Social Networking Misuse (43.5%, bar 4), On-line web security threats (42.9%, bar 12) and Social Engineering Attacks (41.3%, bar 11). The rest of the threats are below 40%: Cyber bullying (39.7%, bar 2), Misuse of internet access exposing devices security (37.5%, bar 8), On-line Sexual Harassment (33.8%, bar 3), Breach of Intellectual Property Rights (33.8%, bar 9) and Unauthorized physical access to corporate facilities (31.4%, bar 13). 9.2% of the participants did not know (bar 14) and 2.3% selected "None of the above" (bar 15).

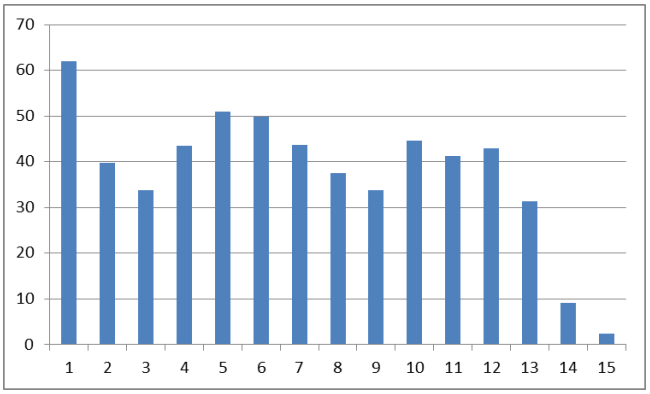


Figure 1. The responses of all participants

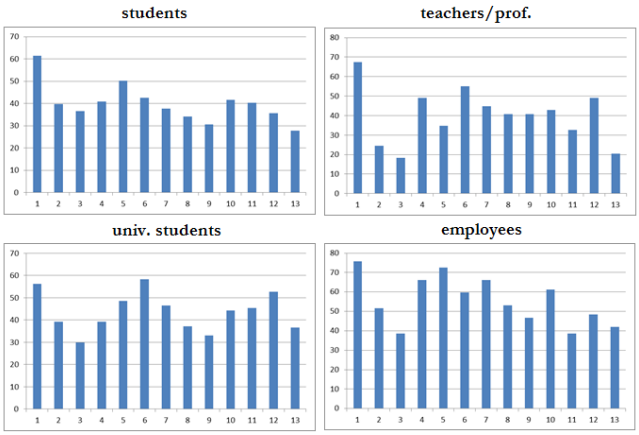


Figure 2. The responses of each target group

Quite similar to the above results are the results obtained when users with previous experience in 3D virtual worlds through a computer were asked to specify the type of security threats that they would like to learn more about: Identity Theft (66.7%) again tops the list, with Cyber Scams (50.5%) and Phishing/Spam (49.5%) following, as also above. The percentages are also quite similar.

The results obtained from each target group vary significantly. In Fig. 2 we present the selections of each target group. The top 3 for each target group are presented next. Students: Identity Theft, Cyber Scams, Phishing/Spam; univ. students: Phishing/Spam, Identity Theft, On line web security threats; teachers/profs: Identity Theft, Phishing/Spam, Social Networking Misuse; employees: Identity Theft, Cyber Scams, Social Networking Misuse. The aforementioned confirm that Identity Theft is very important for all target groups, as it is the first choice of 3 of the 4 target groups and the second choice of the remaining one. In addition, Phishing/Spam is very high in user preferences, while also Cyber Scams and Social Networking Misuse are quite popular.

V. CONCLUSIONS

Besides the specific types of security threats that users would like to learn more about, the most important result obtained from the questionnaire is the confirmation and validation of the following from all of our end user target groups: users believe that 3D virtual worlds could be effectively used for educational purposes by offering educational oriented experiences to the user, users believe that 3D virtual worlds facilitate a 'learning by doing' educational model and users would like to participate in learning sessions facilitated through 3D virtual world simulations.

Moreover, based on the questionnaire data analysis and obtained results presented in Section 4 of this paper, we close this work by describing the most important user requirements, based on which the 3D VWLE, its activities and the security threat scenarios will be implemented:

- Small minorities from all 4 target groups never use a mobile device for online activities. Hence, a security threat scenario that is only about mobile device usage should not be considered.

- 75% of the users are confident in using a computer or a mobile device for online activities. A security threat scenario should validate this information.

- 25% of students do not shop online. For students, security threat scenarios regarding online shopping could be omitted.

- A few of the online shoppers, mostly employees and univ. students shop from anywhere. Security threat scenarios regarding secure online shopping from public places should be implemented.

- Many adult users shop online and manage e-banking accounts but at the same time their confidence regarding e-banking is relatively low in all target groups. Also, very few participants have only an m-banking account, therefore users may not trust mobile devices to manage an m-banking account. E-banking and m-banking security educational scenarios can be developed to educate users on how to safely use them.

- Many users exchange personal and sensitive information over the internet with strangers and/or family members. Employees do it at a much lower percentage than the other 3 target groups. Security threat scenarios about exchanging sensitive information over the internet should be implemented - employees could be excluded from such scenarios.

- 40% of the users do not believe or do not know (mostly students and univ. students) that shopping online may risk the exposure of sensitive data from their side.

Appropriate security threat educational scenarios should be implemented that show how shopping online can be risky and how it can be done securely.

- About 50% of the users believe that their home internet connection is very secure while using an Ethernet cable or a Wi-Fi wireless connection (55% of students). Security threat scenarios regarding home internet security and potential risks should be considered.

- Users in general do not think that using a public computer is very safe but many students stated that they did not know. A security threat scenario regarding public computer usage should be considered, at least for students.

- 25% of the users have stated that they know very well how to protect and secure their electronic data from cyber threats when using a public Wi-Fi. Appropriate security threat scenarios that rate the users through their activities can determine whether this is valid. Also, 20% do not know at all how to do that and 50% don't know exactly how to do it but are aware of some protective measures.

- Security threat scenarios aiming to train users in more "technical issues" are mandatory. The mean percentage of correct answers for all users was only 49%, while only one person responded correct to all questions. Moreover, only 44.8% of the participants responded correct to at least one question (mean value).

- Regarding the technical questions, students have lower success than the other target groups in a few questions and a higher success rate only in one question. Technical oriented security threat scenarios specifically for students (small children) should be considered.

- Half of the users have stated that they have previous experience with computer games - as expected, students and univ. students have the highest rate. Gamification of security threat scenarios for students and univ. students is appropriate and desirable.

- 40% of users have previous experience in 3D virtual worlds through a computer (PC, laptop, mobile device). Students, univ students and employees have the most experience, teachers/profs have the lowest: 50% of teachers/profs do not have previous experience at all. This should be considered while designing 3D virtual world security threat scenarios for teachers/profs. An introductory video/scenario may be needed for this target group before interacting with the 3D virtual world within a real scenario.

- More than half of the users agree that 3D virtual worlds could be effectively used for educational purposes by offering educational oriented experiences to the user, that 3D virtual worlds facilitate a 'learning by doing' educational model and would like to participate in learning sessions facilitated through 3D virtual world simulations. 3D virtual world security threat scenarios for educational purposes should be developed, that will emphasize on a "learning by doing" educational model.

- Regarding the purpose of the users' participation in 3D virtual worlds, 41.4% was Gaming, 19.1% was Educational, etc. 3D virtual world security threat scenarios focused on gaming as well as education are well known to users and should be considered.

- The security threat scenarios to be implemented should include Identity Theft, Cyber Scams and Phishing/Spam. User preferences in the security threat scenarios vary little based on each user's target group.

VI. ACKNOWLEDGMENT

This work has been co-funded by the European Union under the Framework Lifelong Learning Programme / Key Activity 3 - ICT / Multilateral Project (European Commission, EACEA) for the project V-ALERT- "Virtual World for Awareness and Learning on Information Security". This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of its content.

VII. REFERENCES

[1] Virtual World for Awareness and Learning on Information Security (V-ALERT) Official Website, Online: http://v-alert.eu/, 2014.

[2] T. Lombardo, The pursuit of wisdom and the future of education. In *T. C. Mack (Ed.), Creating global strategies for humanity's future*. Bethesda, MD: World Future Society. 2006, 157–176.

[3] Daden Limited. Virtual Worlds for Education and Training. White Paper. 2010. Retrieved February 22, 2011 from http://www.daden.co.uk/media/white-papers.

[4] Daden Limited. Immersive Environments for Learning, Education and Training. White Paper. 2014. Retrieved January 30, 2014 from http://www.daden.co.uk/media/white-papers.

[5] J. Richter, L. Anderson-Inman, and M. Frisbee, Critical engagement of teachers in Second Life: Progress in the SaLamander Project. In *Proceedings of 2007 Second Life Education Workshop*. Chicago, USA, 2007.

[6] M. D. Dickey, Three-Dimensional Virtual Worlds and Distance Learning: Two Case Studies of Active Worlds as a Medium for Distance Education. British Journal of Educational Technology. 36(3), 2005, 439-451.

[7] S. de Freitas, Serious virtual worlds: A scoping study. Joint Information Systems Committee. 2008.

[8] V. Maratou, E. Chatzidaki, and M. Xenos, Enhance learning on software project management through a role-play game in a virtual world. Interactive Learning Environments. DOI: 10.1080/10494820.2014.937345, 2014.

[9] M. J. Callaghan, K. McCusker, J. Lopez Losada, J. G. Harkin, and S. Wilson, Engineering Education Island: Teaching Engineering in Virtual Worlds. ITALICS (Innovation in Teaching And Learning in Information and Computer Sciences), Vol. 8, Issue 3. 2009.

[10] J. Ryoo, A. Techatassanasoontorn, D. Lee, and J. Lothian, Game-based InfoSec Education using OpenSim. In *Proceedings of the 15th Colloquium for Information Systems Security Education Fairborn*, Ohio. 2011.

[11] J. Ryoo, A. Techatassanasoontorn, and D. Lee, Security Education using Second Life. IEEE Security & Privacy. Published by the IEEE Computer Society. (Eds.) Matt Bishop, Cynthia Irvine. 2009.

[12] B. Duffy, Network Defense Training through CyberOps Network Simulations. In *Proceedings of the Modeling, Simulation, and Gaming Student Capstone Conference*. Norfolk, Virginia. 2008.

[13] V. Pastor, P. Díaz, and M. Castro, State-of-the-art simulation systems for information security education, training and awareness. In *Proceedings of the IEEE Engineering Education 2010 – The Future of Global Learning in Engineering Education*. 2010.

Enhancing Learning on Information Security Using 3D Virtual World Learning Environment

Vicky Maratou*, Michalis Xenos*, Divna Vučković**, Andrina Granić***, Aleksandra Drecun**
* Hellenic Open University, Patras, Greece
** Center for the Promotion of Science, Belgrade, Serbia
*** Faculty of Science, University of Split, Split, Croatia
v.maratou@eap.gr, xenos@eap.gr, andrina.granic@pmfst.hr, dvuckovic@cpn.rs, adrecun@cpn.rs

***Abstract*—This paper explores the technology of 3D virtual worlds in relation to their possible benefits for education. The taxonomy of the 3D virtual world platforms (proprietary and open source) are presented and evaluated for the implementation of virtual learning platform of the V-ALERT project. Based on the recent literature, the platforms' characteristics are reviewed. Additionally, the main criteria that are set by the consortium of V-ALERT for the selection of the most appropriate virtual world platform are outlined and the final choice is discussed. The platform of choice is described in more details concerning its functionality and features, as well as its possible integration with other modules such as Learning Management Systems (LMSs), for example LMS Moodle.**

I. Introduction

Numerous 3D Virtual Worlds, formally called Multi-User Virtual Environments (MUVEs), have recently become available, many of which are tuned to specific uses, either for socialization and leisure activities, or for more "serious" purposes such as commercial facilitation (e.g. sales and marketing, or customer support) and education enhancement (e.g. training simulations). The special characteristics and distinct possibilities of the Virtual Worlds (VWs) make them a powerful technological tool towards enhancing the learning experience. This is one of the main reasons for the selection of 3D VW platform for development of the advanced, interactive and motivating tool for rising the awareness on Information Security threats and learning how to recognize and avoid unsafe actions in the scope of V-ALERT project, as teaching and training applications in VWs seem to offer remarkable benefits to students. As stated by [1], "VWs are ideally placed to support pedagogies that aim at moving away from chalk-and-talk learning and focus on more real-world learning styles such as learning through action, cooperation, gaming, problem solving, etc".

The aim of this paper is to present the results of the first phase of the V-ALERT project co-financed by European Commission under the Framework Lifelong Learning Programme / Key Activity 3 – ICT / Multilateral Project [10]. The goal of the V-ALERT project is to support the establishment of Information Security (IS) culture through providing awareness and facilitating learning process using 3D Virtual Worlds platforms. The high proliferation of information and communication technologies (ICT) and everyday use of Internet and computers by majority of people of all age groups for work, learning, entertainment, communication etc. brings a lot of benefits, but also certain risks related to non-informed ICT use. The ICT user should be aware of the basic principles of information security and data protection. This is the reason for the development and implementation of the innovative and immersive e-learning tool in different ICT user target groups (pupils and teachers, ICT students, academics and enterprise employees) in the scope of V-ALERT project. An online 3D Virtual World Learning Environment (VWLE) is being developed which is simulating real-life Information Security threat scenarios, allowing users to gain first-hand experience of different risks and threats, but in a safe manner.

The paper presents the 3D Virtual World platform selection in accordance to project requirements and the transformation of 3D Virtual World to the educational virtual environment for advanced learning on Information Security.

II. Virtual World platforms for the enhancement of the learning process

A. 3D virtual worlds

Although various definitions of the VW have been proposed by different authors, one commonly accepted definition does not yet exist. However, all the definitions have in common the following basic characteristics of the VW:

1. shared space which allows multiple concurrent users to be present,
2. graphical user interface which depicts the virtual environment,
3. immediacy that supports real-time interactions,
4. interactivity that allows users to interact with the virtual environment, providing the means for building, creating and embedding digital content,
5. persistence which ensures that the VW (objects and constructs) as well as any alterations made by the user will continue to exist and function even after the user has left the VW,
6. synchronicity for synchronous users' communication through text and/or voice,
7. network of people who can communicate and interact with each other, forming short term and long term social groups, i.e. a sort of ecosystem,
8. avatar representation in other words a digital representation beyond a simple label or name, that has agency (an ability to perform actions) and is controlled by a human in real time,
9. networked computers managing all data and facilitating the virtual experience.

B. Virtual world platforms

The 3D Virtual Worlds platforms are an innovative ICT technology that provide tools for the creation of highly immersive 3D graphical and interactive online environments which can be either replicas of existing physical places, or imaginary places, or even places that are impossible to visit in real life due to restrictions such as cost or safety. These VW platforms can be either proprietary or open-source.

In the following text the proprietary and open-source VW platforms are presented. Those are currently most popular in the educational community for the development of fully customizable and thematic rich virtual worlds in which multiuser interactive educational simulations, serious games and learning activities can take place. VW platforms that are mainly used in business sector for meetings and collaboration are not included because they would not support the development needs of a project similar to V-ALERT.

C. Proprietery Platforms

The evaluated proprietary 3D Virtual Worlds platforms are Second Life, Active Worlds, Jibe and Unity.

Second Life [12], launched by Linden Lab in 2003, is the most popular of the Social Worlds, with the largest active user and educational community. It features a detailed 3D graphical environment and customizable avatars, built-in voice and standard text communication tools (i.e. chat, IM). SL provides a social network with groups, through which information and object sharing can take place. SL also has an in-world economy (the virtual currency is Linden Dollars (L$): 2500L$=10.09USD) and an enormous market with user-generated virtual goods and tools. One of the most exceptional capabilities of SL is the ability to build objects and write scripts in-world. Registration and basic usage is free but the users have the option of paying a small monthly fee in exchange for a small parcel of land where they can build a home and become "residents". However, serious building projects require the purchase of a private island or a large piece of land (parcel) in the Mainland. A parcel in the Mainland may be made private and accessible only to those who belong to a group, but visitor avatars may still access the neighboring land. Hundreds of learning organizations - from nearly every country- are either augmenting their current curriculum with a virtual learning component or they are holding classes and entire programs exclusively in immersive learning environments in SL. More details on the features of SL are presented in [9].

Active Worlds [13] was launched in 1997 and works much in the same way as SL. Although restricted usage is free through the "tourist" account, paying a small monthly fee allows one to become a "citizen". Only "citizens" can have a unique name, unrestricted access to any part of any world on the platform, avatar customization, object building and access to social networking features such as voice chat, IM and file sharing. For users who need more control over their environment and more privacy, private firewall-protected Universes are available for enterprise and educational projects. These are separate worlds from the main universe and their cost varies. A separate set of worlds and a community for educational projects is also available under the name Active Worlds Educational Universe where over 80 organizations have presence. More details on the features of AW are presented in [9].

Jibe [14] is a multiuser 3D virtual world platform developed by ReactionGrid Inc. The developed virtual worlds can be embedded in any web page or accessed from mobile devices, they can either be hosted by ReactionGrid Inc. or fully installed on private servers. Jibe requires the installation of the Unity web plugin with Android and iOS support under development. Using the Unity 3D editor to build a Jibe virtual world, it results to a professional development environment with professional quality graphics, physics and sound. It allows the creation of 3D objects and the import of 3D models from Maya, Blender, etc. Jibe also features customizable 3D avatars, private/public text chat, user tracking, Vivox voice integration, built-in registration database, integration with Facebook, LMS, CMS, hooks for Augmented Reality apps, support for SCADA and Robotics. More details on some of the features of Jibe are presented in [9].

Unity [15] is not a virtual world platform. It is a 3D (& 2D) professional game development tool which can be used to create suitable training simulations and educational virtual worlds in 3D from scratch which can then be accessed through a client or a web based player. The Unity offers the possibility to develop a game and its user interface without having to program in complex computer languages, such as C++. The language behind the Unity scenes is C#. The development of single-player games/apps requires only downloading and installing Unity but the features and properties of the developed training environment depend mostly on the ability to use the content creation tools. The Unity Asset Store, a global marketplace of objects (as well as code) for Unity, provides content (character models, materials and textures, landscape painting tools, game creating tools, audio effects, music, visual programming solutions, scripts, etc) for a very low cost or even free. Unity evolves with the latest mobile (iOS, Android), desktop (PC, Mac, Linux), Web (web player, Flash) and console (Wii U, PS3, Xbox 360) technology, offering smooth development and deployment of a game with high quality of graphics and solid performance on any device. More details on some of the features of Unity are presented in [9].

D. Open source Platforms

The special emphasis was given to the evaluation of the open source platforms due to the fact that the V-ALERT project is co-founded under the LLP Programme of the EACEA agency of the European Commission which encourages the use of open source software. The platforms OpenSimulator, OpenWonderland and OpenCobalt are compared and analysed.

OpenSimulator [16], often referred to as Opensim, is a free, open-source, 3D application server that allows the creation of 3D virtual worlds, where multiple users can simultaneously be present. These virtual worlds can be accessed through various open source clients and can remain private, behind the firewalls, or become public. OpenSimulator is written in C# and its framework is designed to be easily extensible through external modules. The OpenSimulator project started in early 2007 as an open source server side to Linden Lab's Second Life open-source client. Consequently, OpenSimulator's current architecture is heavily influenced by that of Second Life, allowing the user to produce similar highly detailed 3D graphical environments from scratch at a low cost, or at no cost, provided that the hardware, software

the building, scripting and technical skills are offered for free. The avatars are fully customizable and resemble those of Second Life. The in-world communication is based on text communication tools (i.e. chat, IM). At the moment, a reliable choice for free voice service with lip sync is the one provided by Vivox Inc., by request. An exceptional feature is Hypergrid, a protocol that allows hyperlinking between Opensim worlds and supports seamless avatar transfers among these worlds. Despite the fact that the platform has not reached a beta version yet, it proves to be quite stable and robust. The aforementioned reasons plus the freedom of owning, building and configuring the virtual world, have made OpenSimulator very popular among the educational and science community. Virtual worlds and education. [9] includes some basic features of OpenSimulator.

OpenWonderland [17] is an open source 100% JavaTM toolkit for creating 3D collaborative virtual worlds from scratch. OW is in its early stages of development and although the graphics of the environment are rather simplistic, other features of the platform are comprehensive. The toolkit allows the creation of modules which can extend any part of the system (client or server) and add functionality. Out of the box and with a bit of software development effort, customized, special-purpose virtual worlds can be created. Some examples of the external modules that have been created by different developers and can be downloaded from the Module Warehouse are: Authentication system, webcam viewer, writable (text or HTML) poster, collaborative text editor, etc. Open Wonderland offers the ability to run Java and X11 (Linux) applications inside the virtual world. Almost all Java applications, which can be 2D or 3D, are created with multiple users in mind. For example, there is a shared whiteboard which multiple people can draw on at the same time, there are sticky notes for brainstorming and a multi-user PDF Viewer for browsing slides independently or in sync with a presenter. A distinct feature of Open Wonderland is the ability to easily bring in existing content. The list of document types that can be dragged and dropped into the world is ever growing. Moreover, a creator can import any content found in the Google 3D Warehouse. Open Wonderland does not offer in-world 3D building, but 3D stuff can be imported from Maya, Google SketchUp, Blender, etc. Hence, avatar, instead of being built in-world, must be created on Evolver website and then dragged and dropped into the OW virtual worlds. Also, OW does not support avatar's inventory. Within OW worlds, users can communicate with high-fidelity, immersive audio, share live desktop applications and collaborate in an education or business context (simulations, meeting rooms, mixed-reality worlds, etc). Basic features of OpenWonderland platform are shown in [9].

Open Cobalt [18] is an open source virtual world browser and a toolkit for creating private virtual worlds. OC shares similarities with other 3D virtual environments such as Second Life, but OC uses the peer-to-peer technology instead of servers. Through the OC website, peer-to-peer technology allows its users to access OC virtual worlds on LANs, intranets, or across the Internet without any need to access anyone else's servers. Anyone can host an OC virtual world from all over the Internet for free. Open Cobalt's ability to leverage peer-to-peer technology as a way of supporting interactions within virtual worlds is a major point of difference from commercial multi-user virtual world systems, such as Second Life, where all in-world interactions are managed by central servers. Hence, users can set up virtual spaces and interact with others of their choice with no hosting fees, licensing or virtual land lease costs. Similarly to OpenSimulator, OC makes it possible for people to hyperlink their virtual worlds via 3D portals in order to form a large distributed network of interconnected collaboration spaces. It also offers a set up of public or private 3D virtual workspaces that feature integrated web browsing, voice chat, text chat, and access to remote desktop applications and services. OC lacks 3D content creation tools in-world. It provides the infrastructure for world creation, navigation and collaboration and it supports content created in free or open source authoring applications such as Sketchup or Blender. Through Open Cobalt's VNC capability, web resources (LMS, CMS, wikis, etc.) can be brought into the virtual spaces - interactively. One distinct advantage of OC is the motion simulation. Motion Simulation written in Smalltalk through using FreeCAD application can be easily imported in OC virtual worlds. Basic features of OC are presented in [9].

III. 3D Virtual Worlds and Education

The aforementioned characteristics of the 3D Virtual Worlds could potentially transform these environments to "educational virtual environments". In [6] an "educational virtual environment" is defined as an environment that is based on a certain pedagogical model, incorporates or implies one or more didactic and learning objectives, provides users with experiences they would otherwise not be able to experience in the physical world (or in a classroom) and redounds specific learning outcomes.

Within this context, a rapidly growing interest in learning and teaching within 3D VWs is observed and a large number of schools and universities own virtual spaces for their educational purposes mainly by extending their campuses to the virtual space. 3D educational VWs are usually being used either as safe simulation environments or as virtual classrooms.

The concept of "Encoding Specificity" is a critical issue in the use of simulations for learning. Extended research on human memory, carried out by cognitive scientists and psychologists, show that the "transfer [of learning] is maximum when the conditions at retrieval match those present at encoding." [2]. This means that a learner will be better able to remember what he/she has learned if the conditions during learning match those during recall. In certain educational topics, immersive simulations of the "real life" environment or situation could lead to better recall comparing to only reading books or watching PowerPoint presentations.

In comparison to other e-learning technologies, 3D VWs can provide learners with a full understanding of a situation using immersive 3D experiences which allow the learner to freely wander through the learning environment, explore it, obtain sense of purpose, act, make mistakes, collaborate and communicate with other learners [3]. Indeed, two unique features that the technology of the 3D VWs can offer is the sense of immersion, i.e. the impression of "actually being in there" watching the world through the eyes of the avatar and the sense of presence, i.e. the feeling that the person is an entity of the virtual

world, capable of interacting with other entities in the same way as in a physical space.

The anthropomorphic avatars can enhance the sense of presence [4]. As stated by [5] "an avatar was perceived to be more social present and co-present in the social interaction as compared to an agent represented by a low-anthropomorphic virtual body". Eventually, 3D interactive VWs occupied by human-like avatars could effectively enhance collaborative training activities (games, simulations, and alike). Additionally, immersion accompanied with the interaction with virtual objects can enhance learners' interest and engagement to the learning tasks and help them develop a stronger conceptual understanding, depending on the content [6] [7].

However, according to [4] "it should be considered that the simple use of highly immersive technology alone could not be effective unless it is coupled to specific design strategies, such as for example "goal-based scenario approach". In other words it is important to set training goals for the learner and offer meta-cognition and reflection mechanisms either embedded in the VW or readily available by a trainer who is present.

IV. V-ALERT VIRTUAL LEARNING ENVIRONMENT

The V-ALERT project, co-founded under the LLP Programme of the European Commission, aims to support the establishment of an Information Security culture in different ICT user target groups (pupils and teachers, ICT students, academics and enterprise employees) by providing awareness and training through an innovative and immersive e-learning tool.

A. V-ALERT aims and outcomes

The vision of V-ALERT is to use a uniform environment that is simulating a real-life security threat scenarios, examples and counterexamples in a way that different groups of users are experiencing the risks and combining critical skills, knowledge and collaboration to overcome them, without exposing their organization to real risk. The end-users are immersed in a virtual environment which simulates real-life working or educational environments. These environments are also being enriched with relevant multimedia educational material for further comprehension of the subject.

The underlying pedagogy in this approach is "learning by doing" which can enhance experiential learning, increase intrinsic motivation of learners and lead to deeper understanding and learning. Furthermore, the virtual world learning environment that is being developed in this project is going to be an on-line 24/7 educational tool which can support distant learners of different ages and provide Information Security Awareness in a way that is technologically simple, pleasant, safe and cost effective. The project aims to:

- enrich pupils' awareness on Information Security through collaboration, critical thinking and "learning by doing",
- empower teachers to adopt novel ICT approaches in teaching and also offer them the chance to become aware of Information Security risks,
- help enterprise employees realize their crucial role in securing information assets of the enterprise, allowing them to "safely" experience the security risks, the types of countermeasures available, the situations in which they are suitable and any constraints that they may impose,
- expand and consolidate the existing Information Security knowledge of ICT students and University/ College Professors of various expertise.

An on-line 3D virtual world learning environment (VWLE) is being developed which simulates real-life Information Security threat scenarios, allowing users to gain first-hand experience of the different risks and threats, though in a safe manner (see Figures 1 and 2).

Figure 1: Interactive Slide Presentation "In-world"

Figure 2: "In-world" Library on Information Security Risks

Due to the broad nature of end-users, the implementation of VWLE is reinforcing transversal competences, such as digital competence, while bridging the worlds of education and work. An initial end user requirements collection was conducted through questionnaire, interviews and focus groups, while participated users from Cyprus, Greece, Serbia, Croatia and Bulgaria have been categorized in 4 different target groups: (i) students of primary or secondary education, (ii) teachers or academic professors, (iii) ICT college or university students and (iv) enterprise staff or employees in an organisation or administrative personnel [11].

Besides the specific types of security threats that users would like to learn more about, requirements gathering revealed the following:

- users believe that 3D virtual worlds could be effectively used for educational purposes by offering educational oriented experiences to the user,
- users believe that 3D virtual worlds facilitate a "learning by doing" educational model, and
- users would like to participate in learning sessions facilitated through 3D virtual world simulations.

The virtual environment is providing scenarios in a core language, but is easily expandable in any new language. Within the scope of the project, the in-world learning material and scenarios will be available in English (as a

core language) as well as in Greek, Serbian, Croatian and Bulgarian. Furthermore, V-ALERT main outputs are listed in the following:

- conceptual design of Information Security threat scenarios concerning the needs of each target group,
- customizable on-line 3D virtual world designed to best implement these scenarios through in-world engaging learning activities and simulations,
- supportive documentation and Information Security educational material for the end-users,
- reports on the formative evaluation of the virtual world learning environment and the pilot implementation of this environment in each partner country.

B. V-ALERT platform selection

The VW platform selection criteria were set in line with the aims of the V-ALERT project, the development requirements, the context philosophy and the budget. General criteria is presented in the following list:

1. cost-free and open-source,
2. development of fully customizable and multiuser virtual worlds in order to simulate various scenarios,
3. good system stability,
4. straightforward server configuration and parameterization in order to fully control the VW and the usage rights at will,
5. self-hosting possibility,
6. reasonable hardware and bandwidth requirements that can be provided by Hellenic Open University's Software Quality Lab,
7. user-friendly and free downloadable, multi-platform client software allowing for non-Windows users' participation,
8. high-quality 3D graphics and human-like fully customizable avatars to support the issues of immersion and presence,
9. platform popularity for educational projects along with large, active and supportive community of developers.

Furthermore, a number of specific criteria was also taken into account:

1. built-in 3D editor for in-world creation and editing of 3D virtual objects and landscapes,
2. expandable functionality and in-world interactivity through scripting language,
3. possibility of free object/landscape/virtual world import and backup,
4. real-time communication through text chat, IM and Voice,
5. in-world scripting language to support the implementation of NPC creation and programming,
6. possibility to embed LMS/VLE functionality inside the virtual world.

Additional to the aforementioned required general and specific criteria, the existence and availability of free, rich, open, and customizable pre-made content (i.e. 3D objects, scripts, functionality modules) was considered a benefit.

Based on the first general criterion, the platform of choice must be open-source and free. Therefore, comparing the features and the functionality of the three open-source platforms that are presented in Chapter II of this paper, only OpenSimulator is more mature and full-fills all the requirements described. Additionally, OS's compatibility to SL, the most popular 3D virtual world platform for educators all over the world, as well as its open and modular design makes OS platform ideal for educational institutions and enterprises that need to have full control and maximum flexibility on their 3D simulations, in virtual worlds that offer pretty much the same graphics, functionality and building possibilities as Second Life but in significantly lower cost (or at no cost at all). Furthermore, anyone who wishes to develop virtual worlds on OS can benefit from the intellectual outcomes of two very large and active educators' and developers' communities, which is a major advantage for someone who is not experienced in this field.

Although OW and OC platforms present some quite interesting features, they are still in earlier stages of development and evolve in slower rate comparing to OS. They are supported by smaller community of developers, especially the OC platform. For the V-ALERT project and its requirements, the fact that OW and OC require the use of external 3D modelling applications for avatar customization and 3D object creation and editing as well as the fact that they do not offer a built-in mechanism for NPC creation and programming, was considered as drawback that could increase the total workload of the development. Finally, the LMS/VLE functionality, that can facilitate the accomplishment of the project's objectives, can be integrated only to the OS platform through the open-source SLOODLE [19] module of Moodle LMS [20].

C. OpenSimulator as the basis of V-ALERT platform

The 3D VWLE is being developed on the open source VW platform OpenSimulator and is integrating functionality of the Moodle LMS through SLOODLE middleware.

OpenSimulator 3D VW platform (current version 0.8.) facilitates the development of 3D environments using a variety of technologies. It is easily extendable through loadable modules that enable to build completely custom configurations and embed extra functionality to the platform. OS is being developed in C# programming language, and additional functionalities can are added using loadable modules. OS is released under a BSD License, making it both open source and commercially friendly to embed in products. Out of the box, OS is used to simulate a virtual environment similar to SL (including client compatibility due to the same client-server communication protocol). Although OS is still considered alpha software version, its improvement is progressing rapidly and at the moment the platform is considered relatively stable and robust to be used for the development of rich and immersive multiuser virtual worlds.

At the broadest architectural level the main components of OS system are three: a) the server (or "simulator"), b) the client (or "viewer") and c) the services [8].

The OS server is responsible for the maintenance and update of the virtual world status, managing every user and/or object applied alteration to the virtual environment state. For this reason it is also called "simulator". The client or "viewer" is the software responsible for the 3D graphical rendering of the avatars and the virtual world and acts mainly as an interface between the user and the simulator(s). The backend of the system consists of the

Services which provide the virtual world simulator(s) with the common resources requested. An example of the interaction between client, regions simulator and services in classic standalone architecture is shown in Figure 3. As presented in the figure, only the simulators have access to the VW services – clients, except for login phase, always send and receive data through a simulator instance.

OpenSimulator can operate in two different modes: a) Standalone or b) Grid mode. In standalone mode, a single computer process (i.e. OpenSim.exe) handles the entire simulation and the services of the virtual world. Standalone mode is simpler to configure, but is recommended for "light" worlds and smaller number of concurrent avatars. In grid mode, the services are not part of the VW simulation process (OpenSim.exe). Instead, they run in a separate process (Robust.exe).

The main advantage of using virtual worlds is, according to most authors, the impressive 3D graphics that create an environment that can attract user's attention and facilitate immersion. Built into the software of OS is a 3D modeling tool which is based around simple "prims", allowing the avatars to link them and build complex virtual objects.

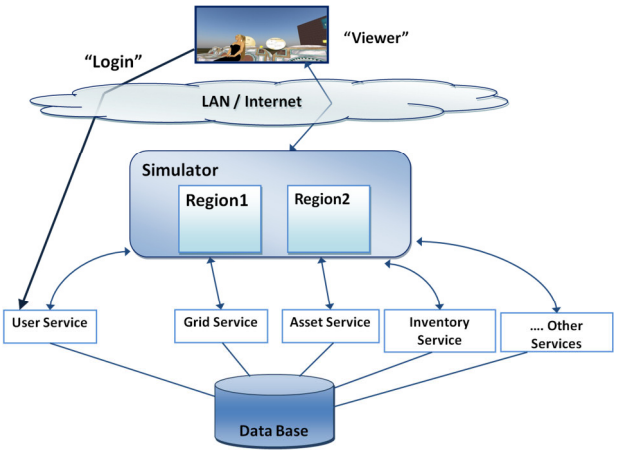


Figure 3: OpenSimulator architecture

V. Concluding Remarks

Based on the literature survey about the 3D virtual world platforms used for educational purposes and the aims and objectives of the V-ALERT project, it is concluded that the most appropriate platform is the OpenSimulator. The OpenSimulator offers all the functionality needed for the development of the threat case simulations, it is expandable and open-source. On the other hand, it can be integrated with Moodle LMS through the SLOODLE middleware. This additional functionality is going to be extremely helpful for the enrichment of the virtual world simulations with educational materials related to Information Security and it can also offer the option of keeping record of the users' in-world performance and profile. The latter option will aid the application of the recommendation algorithms on the users' stored data in order to provide real time in-world personalized recommendations.

Acknowledgment

This work has been carried out within the project V-ALERT - "Virtual World for Awareness and Learning on Information Security", partially supported by the European Commission under the Framework Lifelong Learning Programme / Key Activity 3 – ICT / Multilateral Project. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of its content.

References

[1] Daden Limited. (2010). Virtual Worlds for Education and Training. White Paper. Retrieved February 22, 2011 from http://www.daden.co.uk/media/white-papers.

[2] Clark, R.C., & Mayer, R.E. (2003). E-learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multi-media Learning; John Wiley & Sons.

[3] Daden Limited. (2014). Immersive Environments for Learning, Education and Training. White Paper. Retrieved January 30, 2014 from http://www.daden.co.uk/media/white-papers.

[4] Mantovani, F., & Castelnuovo, G. (2003). Sense of Presence in Virtual Training: Enhancing Skills Acquisition and Transfer of Knowledge through Learning Experience in Virtual Environments. In G. Riva, F. Davide, & W.A IJsselsteijn (Eds), *Being There: Concepts, effects and measurement of user presence in synthetic environments* (pp. 167-181). Amsterdam, The Netherlands: Ios Press.

[5] Nowak, K., & Biocca, F. (2001). The influence of Virtual Bodies and Agency on Co-presence, Social Presence and Physical presence. In Proceedings of 2001 of the 4th Annual International Workshop PRESENCE 2001. Temple University, Philadelphia, PA, USA.

[6] Richter, J., Anderson-Inman, L., & Frisbee, M. (2007). Critical engagement of teachers in Second Life: Progress in the SaLamander Project. In Proceedings of 2007 Second Life Education Workshop (pp. 19-26). Chicago, USA. Retrieved October 20, 2012 from http://www.garito.it/prog/psico08/testi-def/slccedu07proceedings.pdf.

[7] Dickey, M.D. (2005). Three-Dimensional Virtual Worlds and Distance Learning: Two Case Studies of Active Worlds as a Medium for Distance Education. *British Journal of Educational Technology*, 36(3), 439-451.

[8] Clark-Casey, J. (2010). Scaling OpenSimulator: An examination of possible architectures for an Internet-scale Virtual Environment Network. Dissertation for the MSc degree in Software Engineering, Kellogg College, University of Oxford.

[9] D2.2 Report on 3D virtual worlds platforms and technologies, V-ALERT project 543224-LLP-1-2013-1-GR-KA3-KA3MP

[10] Virtual World for Awareness and Learning on Information Security (V-ALERT) Official Website, Online: http://v-alert.eu/, 2014.

[11] Mettouris, C. *et al.* (2015). Information Security Awareness through a Virtual World: An end-user requirements analysis, ICIST 2015, (paper submitted for the review)

[12] www.secondlife.com

[13] www.activeworlds.com

[14] reactiongrid.myshopify.com

[15] unity3d.com/unity

[16] www.opensimulator.org

[17] www.openwonderland.org

[18] www.opencobalt.org

[19] http://www.sloodle.org/docs/

[20] https://moodle.org/

A Flexible, Process-Aware Contract Management System

Miroslav Zarić*, Zoran Miškov**, Goran Sladić*
* University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia
** PD Elektrovojvodina doo, Novi Sad, Serbia
miroslavzaric@uns.ac.rs, Zoran.Miskov@ev.rs, sladicg@uns.ac.rs

***Abstract* — this paper presents an open source based, general purpose contract management system. Contract management systems are becoming increasingly important component of business application landscape, aimed at improving the management of complete contract lifecycle. Contract management systems may be viewed as specialized document management systems (DMS). Although such systems can exist and be functional without implemented business workflow, for contract management this process perspective is quite important in order to have time-efficient, yet reliable contract handling. Hence, most of the contract management systems tend to deploy, or even impose, some business workflow. The system described in this paper is developing around an open source business process management system at its core, allowing for flexible contracts management workflows, and creating an extensible platform for customization to different needs.**

I. INTRODUCTION

In recent years, one of the trends in development of business applications is a focus on electronic document management (EDM) systems. EDM system deals with management of documents [1]. In business entities, different documents are often created as a result of some activities in business process or are an essential requirement for fulfilling some specific task. During the business process execution, different individuals undertake activities according to a specified sequence prescribed by some workflow. Workflow Management Coalition defines workflow as the computerized facilitation or automation of a business process, in whole or in part [2]. EDM are usually implemented as document-centric workflow management systems [3].

Any business entity has some contract management procedures in place, often involving a large amount of manual processing. In today's modern and dynamic markets, there is a growing need for time-efficient contract creation and later contract lifecycle management.

Usual lifecycle of the contract involves several phases. In initial phases, legal department is drafting a contract, making sure that the content of the paragraphs is legally clear. After this verification, contract proposal is negotiated with the client, subject to changes. Each change also needs a seal of approval from the legal department. After the contract is negotiated and comes into force, devoted contract management requires further tracking of contract implementation and contract expiration.

However, the need for prompt contract creation is often confronted by the even greater need for legal clarity and content disambiguation in order to protect business entities from wrongful legal claims. Due to time-constraints, manual processing is becoming less acceptable, and prone to errors, resulting in prolonged contract negotiations. Defining the specific contract creation and contract lifecycle management workflows can be beneficial for overall business activity of an entity. IT support for such lifecycle gave rise to specialized contract management systems. In essence, they represent document management system supported by some well-defined workflow.

Contract management is a sensitive issue, since contract content can directly bring benefits for a business entity but can also expose it to unwanted circumstances. Although standard procedures for contract creation and approval exist in any business entity, creating applicative support for them is challenging. Large companies can have enough IT resources to develop customized solutions. Most in-house solutions are created for specific contract types and don't cope well with contract type changes. Medium size and small companies are usually better served by some existing products, but then the issue of reconfigurability of chosen solution may arise.

There are different tasks contract management systems should perform:

- Contract drafting and templates creation
- Contract negotiation and approval, version control
- Milestone tracking and compliance management
- Notifications
- Completion tracking
- Document search and retrieval

Although such systems can exist and be functional without implemented business workflow, this process perspective is quite important for contract management in order to have time-efficient, flexible, yet reliable contract handling. Hence, most of the contract management systems tend to deploy, or even impose, some business workflow. The system presented here has also started as an application tailored to specific needs, but it soon become obvious that more general and process-aware approach is needed in order to create robust and flexible system.

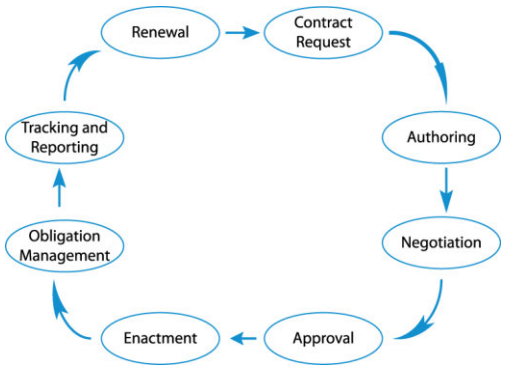


Figure 1. Typical phases of contract lifecycle

Although there are a few slightly different definitions of contract lifecycle phases, most of them agree that full contract lifecycle consists of common phases depicted in Figure 1.

There are different solutions available commercially as a standalone products (Prodagio, Agiloft Contract Management Software, Selectica, Blueridge Software Contract Assistant…) or available as a part of a larger business suite (Microsoft Dynamics CRM, SAP CRM, Oracle Contract Lifecycle Management module…) as well as some freeware solutions. Some solutions are installable applications, some are web-based application deployable on customer servers, and some are offered on a "Software as a Service" (SaaS) model. SaaS model for contract management systems has its set of challenges as described in [4].

The development of the system presented in this paper is centered on open source business process management system Activiti [5]. This approach allows for flexible contracts management workflows while implemented document model is allowing for general purpose implementation.

The aim of the system is to support contract management throughout the whole process:

- By allowing easy contract drafting, with a repository of preapproved general purpose contract paragraphs; drafted document are also subject to approval of the legal department (resolving any outstanding issues which may arouse from conflicting paragraphs
- By allowing contract drafts negotiations with a client
- Final approval and signing of the contract
- Prompt notification of expiring contracts, allowing for contract renewal (especially important for service providing companies such as telecom operators, utility companies, and similar)
- Thanks to the process engine at its core workflow is easily adjustable to different needs
- The core of the system is built as a service-oriented application, allowing for further expansions and client application development
- One client implementation is a web application.

II. Contract Templates And Contract Document Model

Contract management system is, in its nature, document oriented. International standard ISO 82045-1 [6] defines principles and methods of document management. By the given definition, to facilitate document management within their life cycle and for their exchange between partners, documents shall be associated with a set of metadata, i.e. data identifying and/or describing the document [6]. In addition, we can distinguish single document (document associated with its metadata), compound document (composed of more than one document types) and associated with its metadata, document aggregation (assembly of self-contained documents), document sets, and linked documents. Version control is also one of the important issues that need proper consideration.

When contracts are taken into consideration, it is important to notice that they consist of variable data, and in most cases, fixed legal paragraphs. Variable data related to contractual parties and contract terms (such as contract validity period, special discounts granted etc.), may be embedded in paragraph text. Variable data is associated with paragraphs. Since VariableModel allows definition of variable types, they can be of any type relevant to the contract context. One important use of variables is to allow references to other paragraphs.

One approach to contract creation may be to use document templates. In manual processing, usually a Word documents are saved as templates, and manually filled with variable data during contract creation. This manual editing is error prone.

Furthermore, having a complete document template as a basic unit of work turned out not to be very flexible, since it may require a major redaction if terms of a contract is to be adjusted. For this reason, we adopted the view of document template as a sequence of paragraphs.

For contract creation, it is beneficial if these paragraphs are preapproved. That allows for new contract templates to be easily created by assembling existing paragraphs, which can be easily added, reordered, or removed. Even though templates are compiled from preapproved paragraphs, they also need final approval, since paragraphs can be internally correct, but their correlation with other paragraphs may not be acceptable. When template is approved, it can be used for fast creation of valid contracts. Usually, for most business-to-customer relations this is sufficient, and during contract preparation, only variable parts need to be filled.

However, for more complex contracts (usually business-to-business), a negotiation phase of contract lifecycle usually produces slight changes to the original text, therefore, and in this case, flexibility is needed. We added a degree of flexibility by specifying if some paragraph texts are editable. Hence, when new contract is prepared from the template, template parts are instantiated, and editable paragraphs are unlocked. Nevertheless, such an option requires that the final

contract goes through final proof-reading by legal department before final approval by contractual parties.

Basic conceptual model of contract document is given in figure 2.

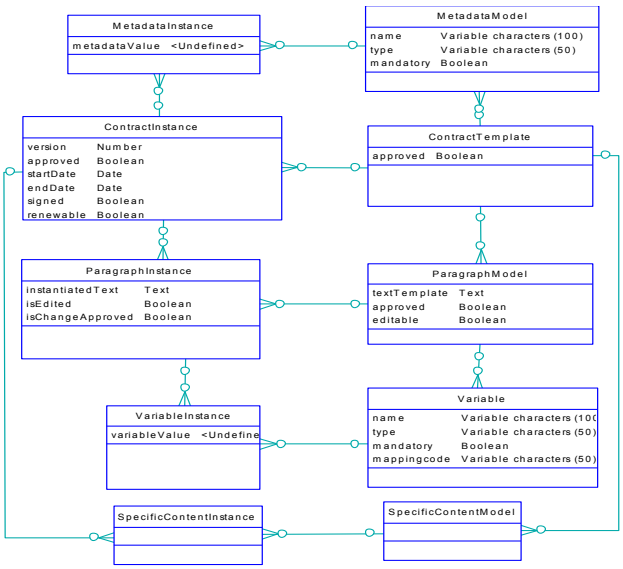


Figure 2. Conceptual model of contract document

This document model has proved to be sufficient for most purposes, especially for a creation of relatively simple contracts. However, for more complex documents, for instance, telecom operator contracts or some utility company contracts, with a vast number of tariff modules, options, and packages, such simplistic approach is not productive. In order to support complex data we added an extension point to the model, represented by SpecificContentModel entity. In the implementation, this entity is interpreted as an abstract class. Specific implementations can then inherit and create their data model to suit their specific needs. One of the specialized implementation was developed for telecom operator representative agents, responsible for contracts with individual users and small companies. In this case, contract needed to have a representational framework for offered packages, tariffs, tariff options and granted discounts.

III. Modelling Contract Lifecycle Management Process

As stated earlier, contract management is a process. Although other approaches, like [7], are viable, the lack of processing support (process logic entangled in the system code) becomes a problem when support for different types of contracts or even slightly different contract management workflows needs to be added to the system. Use of some business process management (BPM) system to manage this process adds a degree of flexibility to the system. In this case, contract management workflow can be expressed as a process model in some process modelling language (BPMN, BPEL, XPDL…). Languages vary in degree of their representational completeness and their applicability to some process engine. With enough technical details entered into the model, a process engine is capable of running the process according to the "prescription" given by the process model. More details about concepts, languages and architecture of process modelling systems may be found in [8]. Some analysis of different languages is given in [9, 10]. In last few years BPMN language [11], developed by Object Management Group, has become a widely accepted standard.

Our implementation is relying on open source Activiti process engine [5]. This engine is also used by one of the leading open source document management systems – Alfresco, to add process support. Although using Alfresco with defined workflow was one solution to contract management, we wanted our own flexible document model and extension points. The choice of Activiti was driven by our prior experience with BPM systems [12], its status of open source solution, and simple API.

BPMN defines Flow Objects (events, activities, gateways), Artefacts (Data Object, Annotations), Connecting Objects (Sequence flow, Association, Message Flow), and Swimlanes as building blocks for creating a process model.

For successful process modeling, it is important to observe functional perspective of process (what needs to be done) and organizational perspective (who is responsible for activities). Also important are informational perspective (what information process execution needs), as well as IT landscape perspective, which defines a correlation to existing IT systems. Organizational perspective is in BPMN modelled through users, groups and user membership to groups. In BPMN notation, organizational aspects are depicted by Pools and Swimlanes. Existence of these symbols in a process model is a visual aid, since it doesn't have a semantic meaning for execution, i.e. activity placed in one Swimlane still needs to be explicitly assigned to a specific role.

In our system, we have implemented a basic contract management workflow suitable for most general purpose requirements. This workflow is given in Figure 3. In this process model, Contract Authoring is a complex activity that can be executed as a standalone process, and is accordingly represented as Call Activity. Similar applies to Contract implementation monitoring.

Since small business often don't require special procedures for Contract implementation monitoring, it is in a conditional branch of the process model. If monitoring is required upon contract approval, it will be executed, presumably for the whole period of contract duration. If so, a separate monitoring process should be modeled taking into account all activities required for contract context, and this model should be provided. Otherwise, process is captured by internal timer set to the contract expiration date, hence keeping the process and contract in an active state. If further actions after contract signing are not required, this timer can be set to a current time (or model changed to be skipped altogether). For

renewable contracts, a message activity is used to send notification that a contract is ending and initiate contract renewal process. Figure 4 shows the model of the contract authoring process.

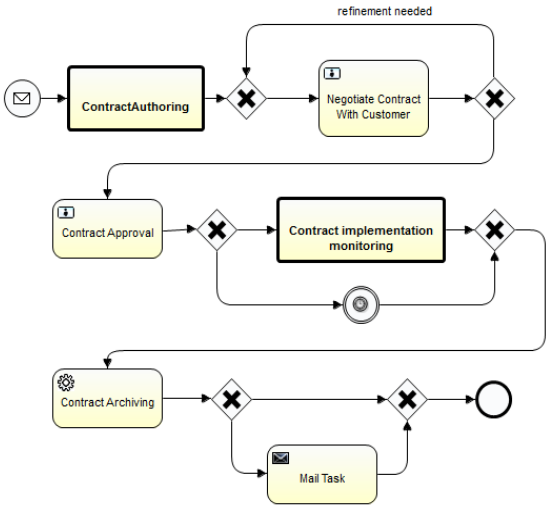


Figure 3. Basic contract management workflow

After request for contract is received, sales staff will determine appropriate contract type and check if prepared template for such contract exists. If it does, it will be used. If appropriate contract does not exist, user will be able to create a new one from existing preapproved paragraphs (possible extension is to allow free text paragraphs so that users are allowed to create new body of text, but this option should be controlled by access control rules). Such contract template proposal now has to go through approval process. After the user has got hold on approved contract template (previously existing, or newly created), it fills in contract data. During this process, if other things, rather than just variable data, are changed (text of paragraphs), the contract will be marked as "dirty" and will require further approval.

In order to evaluate if proposed model is adaptable to different needs, we've modeled a process of contract management in electricity supplier (utility company). Studying their contract management activity diagrams, we have learned that, apart from more elaborate definitions of user roles, and more detailed activities or their changed order, there are no significant differences. Hence, main modelling activities in this case were concentrated on an organizational perspective, i.e. mapping organizational hierarchy of power utility company to user and groups in identity module. Also, a significant attention was given to information perspective – since the process has well-defined support documents and information objects that are used to fulfill some tasks. This detailed process is depicted in Figure 5. If an electronic signature is used on documents, then automatic mailing tasks may be used for final document delivery to the client. Otherwise, the final task needs to be performed as a standard user task. A system task (service task) is used to automatically transfer client data into a client database.

IV. System Architecture

At the core of our system, as mentioned before, is an Activiti BPM engine. A relational database, in our case MySQL, is used by the engine, and accordingly, by the rest of our system. That is configurable, since BPM engine can use any database supported by the Hybernate ORM system. Since Activiti is implemented in Java programming language, and due to our prior experience in development of enterprise applications in Java, we have opted to develop contract management web application also in Java. Such a choice was not mandatory nor forced upon us by the Activiti BPM engine as it exposes most of its API through REST services. Therefore, a developer can choose other options to access BPM engine. In that case, some consideration is needed if event listeners or service classes are specified in the process model. Activiti is implemented as a standard Java package and may be used as a standalone engine, or in other environments, such as a servlet or EJB containers. Since our primary implementation was intended to be a web application,

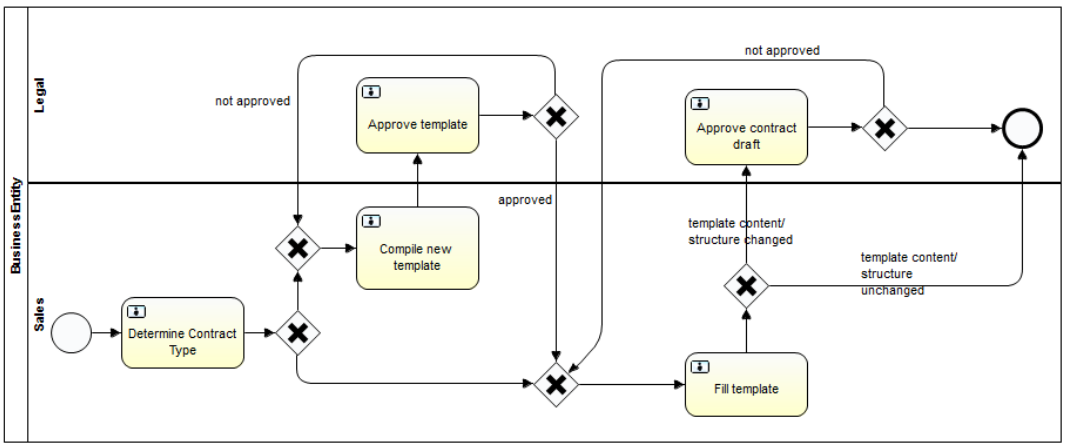


Figure 4. Contract authoring workflow

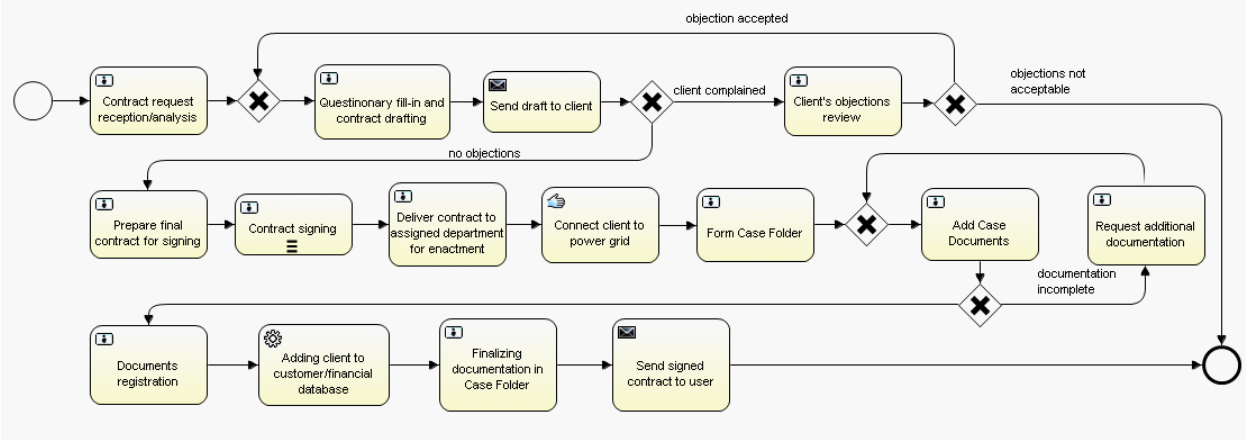


Figure 5. Contract workflow for Electricity Company

Activiti engine is also deployed on the application server. Client application, in our case web browser, is accessing the application via HTTP(s) protocol. However, since functionalities of the application may be exposed through REST services, a different type of client applications can be developed. That option is very useful when an integration of contract management into other business applications is required. If that is the case, a contract management application will use existing services, but can be implemented as a module in larger software system.

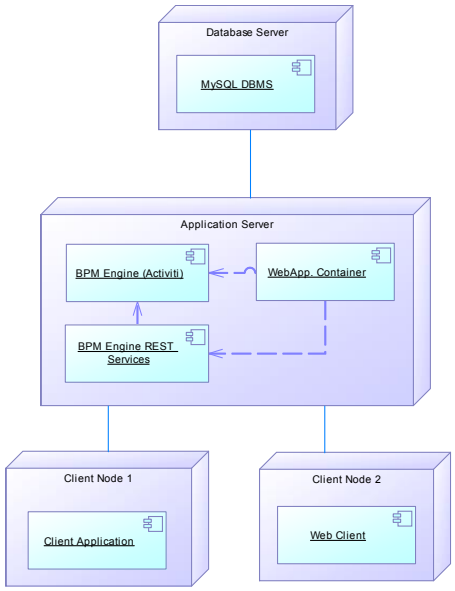


Figure 6. Deployment diagram of application modules

Activiti engine FormService features give us an opportunity to dynamically create user interface for specific tasks. If this feature is used, specific data needed to perform certain tasks may be declared in the model. Simply by reading the model, the process engine is capable of building necessary forms for users. However, if the data necessary to perform the task is dependent on contract type being processed or data created during the contract management process, there is a high probability that we don't have enough information, at the time of modelling, to specify form properties. However, developers can create customized forms, or even generate them dynamically, and attach them to the tasks through the form key property. This approach can be used to map our specific data to user tasks in the process.

The deployment diagram of the core modules is given in Figure 6.

V. Addressing Security Issues

Contract management is a sensitive issue, hence requiring appropriate measures to enhance contract confidentiality and access control is mandatory. Our pilot implementation is developed as a web application. In a case of in-house installation, it is a good practice to allow access only from LAN. Additionally, in order to enhance security, it is recommended that server setup enforces access through HTTPS channels. It is especially important in case of installation on publicly accessible servers. Access control is based upon Activiti BPM engine Identity module, using role-based access control. Most of the data necessary for process running is controlled by the process engine and Hybernate ORM mapper, using parametrized queries, thus lowering the possibility of SQL injection attacks. Data coming from input forms are also preprocessed for this threat. Strings that comes from user input (HTML forms in our case) are checked for XSS (Cross Site Scripting) attacks. Also, appropriate measures, to prevent CSRF (Cross Site Request Forgery) attacks, are implemented in the server application. Database servers are installed on a separate node (server), and access is allowed only through application server, with appropriate credentials.

VI. Conclusion

In this paper, a contract management system, as a specific type of document management system, is presented. A document model, suitable for general purpose contracts, but extensible for complex contract

structures is also presented. The presented system is developed using BPM engine at its core. Basic contract lifecycle management business process has been described, modelled and implemented in pilot project. Additionally, in order to assess overall functionality of the system, several real-life contract handling business processes has been modelled and implemented.

Further work on this system will be concentrated on extending access control features with COBAC (*context-sensitive access control model for business processes*) [13]. Although presented document model has been sufficient for the current implementation, it will also be enhanced by extending paragraph properties, making it possible to express more complex relationships between parts of a document. Such enhancement will bring us an opportunity to develop a system that will keep track of compatible and incompatible document parts, as designated by legal department. In that case, template creation would be simplified, and user would be warned if some parts of a document are conflicting. Additionally, since new standard "The Open Contracting Data Standard" [14] is emerging, aimed primarily at government contracting procedures, such as public procurement, this data model also needs to be taken into consideration.

Multi-tenancy, i.e. ability to offer this system as a service to multiple business entities will be further explored. In that case, special attention must be given to a strict separation of content, access rights, and also possible encryption of data may be appropriate.

REFERENCES

[1] International Organization for Standardization (ISO). (2001): "ISO IEC 82045-1: Document Management – Part 1: Principles and Methods", ISO, Geneva, Switzerland.

[2] Hollingsworth, David: "The Workflow Reference Model", Workflow Management Coalition, Cohasset, MA, USA (1995).

[3] Krishnan, Rupa, Lalitha Munaga, and Kamalakar Karlapalem: "XDoC-WFMS: A Framework for Document Centric Workflow Management System". *In H. Arisawa, Y. Kambayashi, V. Kumar, H. Mayr, and I. Hunt (eds): Lecture Notes in Computer Science 2465*, Berlin: Springer, (2002). pp. 348-362.

[4] Kwok, Thomas, Thao Nguyen, and Linh Lam. "A software as a service with multi-tenancy support for an electronic contract management application." *Services Computing, 2008. SCC'08. IEEE International Conference on. Vol. 2. IEEE*, 2008.

[5] Activiti BPM Platform, http://www.activiti.org/, accessed: January 2015.

[6] International Electrotechnical Commission, "Document management - Part 1: Principles and Methods", *INTERNATIONAL STANDARD ISO-IEC 82045-1,* 2001.

[7] Wenwen, Ding, Chen Yan, and Jiang Zhuoren. "Design of Contract Management System Based on J2EE Architecture." *Business Computing and Global Informatization (BCGIN), 2012 Second International Conference on*. IEEE, 2012.

[8] Mathias Weske: "Business Process Management: Concepts, Languages, Architectures", 2nd ed. 2012, XV, 403 p. 300 illus. Springer-Verlag Berlin Heidelberg 2012, Springer Link

[9] List, Beate, and Birgit Korherr. "An evaluation of conceptual business process modelling languages." *Proceedings of the 2006 ACM symposium on Applied computing*. ACM, 2006.

[10] Indulska, Marta, et al. "Representational deficiency of process modelling languages: measures and implications." (2008). In Golden, W. and Acton, T. and Conboy, K. and van der Heijden, H. and Tuunainen, V., Eds. Proceedings 16th European Conference on Information Systems, Galway, Ireland.

[11] Business Process Model and Notation, Specification, Object Modelling Group, http://www.bpmn.org/, visited January 2015.

[12] Zaric Miroslav, Segedinac Milan, Sladic Goran, Konjovic Zora, "A Flexible System for Request Processing in Government Institutions", *ACTA POLYTECHNICA HUNGARICA, vol. 11, br. 6, p. 207-227,* 2014.

[13] Gostojić, Stevan, Goran Sladić, Branko Milosavljević, and Zora Konjović: Context-sensitive Access Control Model for Government Services, Journal of Organizational Computing and Electronic Commerce, vol. 22 no. 2, (2012) pp. 184-213

[14] Open Contracting Partnership, "The Open Contracting Data Standard", http://standard.open-contracting.org/, accessed: January, 2015.

Digital Technologies for Cultural Heritage Presentation in Bosnia and Herzegovina

Selma Rizvić*
* Faculty of Electrical Engineering Sarajevo, Bosnia and Herzegovina
srizvic@etf.unsa.ba

***Abstract*— For centuries Bosnia and Herzegovina has been a country where East meets the West. Therefore it is very rich with remains of various cultures, nations and religions. Some of these remains are in very bad condition, some have completely disappeared. Digital technologies offer a great potential in preservation and presentation of cultural heritage, enabling better understanding of the common past. This paper provides an overview of 10 years long work of researchers gathered around the Sarajevo Graphics Group and through their projects offers an insight into various applications of Internet and IC technology as a new media for cultural heritage presentation.**

I. INTRODUCTION

Illyrian, Roman and Slav tribes have inhabited this region ever since the Neolithic period. Remains of human settlements were found in Butmir near Sarajevo. Bosnia as a country was first mentioned in the tenth century AD, in the work of the Byzantine emperor and writer Constantine Porfirogenetus. In the medieval European circles Bosnia was extremely appreciated: it had a royal family, palaces, strong and powerful nobility and a unique culture. Ottoman and Austrian Hungarian occupations left traces of their cultures in architecture, infrastructure and lives of people.

The Ottoman Empire offered shelter to the Jews who fled the Spanish and Portuguese Inquisitions. Many Jews settled at that time in the Ottoman province of Bosnia and Herzegovina. Hence the original Jewish community of Sarajevo was Sephardic, and Bosnia hosted Europe's largest Sephardic Jewish community after Spain.

The rich and turbulent history of Bosnia and Herzegovina left us with many cultural heritage objects and sites. Preservation and presentation of this heritage will enable us to better understand our past. Digital technologies offer a new media for exploration and analysis of cultural heritage. They finally made possible the travel through time. Sarajevo Graphics Group is experimenting with application of these technologies on Bosnian cultural heritage.

Ten years ago the researchers from the Faculty of Electrical Engineering Sarajevo were supported by UNESCO in digitization of the most important Bosnian medieval gravestone, stećak from Donja Zgošća, using laser scanner. It was the beginning of a serious research in digital cultural heritage field in Bosnia and Herzegovina. Today the group includes members not only from ICT, but also archaeologists, digital artists, historians, writers and museum curators, as only a true interdisciplinary team can implement these projects with success.

This paper will offer an overview of projects that have been implemented by Sarajevo Graphics Group, through technologies they used and experience gathered in the process.

II. CULTURAL HERITAGE DIGITIZATION WORKFLOW

Although the cultural heritage digitization projects differ in application purpose and technologies used, we define a typical project creation workflow, based on our experience (Figure 1).

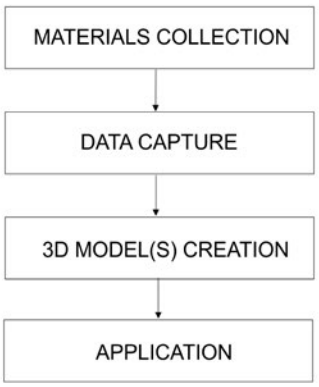


Figure 1. Cultural heritage digitization workflow

Every project starts with collecting materials from historians, archaeologists and cultural heritage protection institutions. Sometimes the information on the objects is obtained from oral histories or memories of people. This step is very important for the success of the project.

Data capture could be performed using various techniques, such as laser scanning, LIDAR (Light Detection and Ranging) data, photogrammetry etc. It is important to note that data captured in this stage needs a lot of post-processing in order to be used in 3D model creation.

The 3D model of a cultural heritage object could be created using classical modeling techniques, mostly in case when no remains of the object are preserved, but it can be also generated using a combination of geometry obtained in the data capturing stage with newly created geometry.

The last step in the workflow depends on the purpose of the cultural heritage digitization project. The 3D model can be textured and rendered to be a part of a digital story, or could be exported in web 3D format and used as interactive virtual environment. There are also models generated for preservation purposes, for education or reproduction using 3D print.

In the following sections we will present our projects within the described workflow and elaborate the digital technologies we used in the implementation.

III. DATA CAPTURE

Cultural heritage objects with irregular shape, containing carvings or other decorations, can not be easily modeled using classic modeling techniques. Laser scanning offers a great support in capturing all details of object's shape and surface.

Our first encounter with laser scanning was during the UNESCO funded project "Virtual reconstruction of cultural heritage in Bosnia and Herzegovina" in capturing the famous Bosnian medieval gravestone, presently preserved in the Botanical garden of the National Museum of Bosnia and Herzegovina. We used Konica Minolta 910 laser scanner and performed a number of individual scans of surface sections, which were later connected to a single geometric mesh (Figure 2).

Figure 2. Laser scanning of stećak from Donja Zgošća

The obtained data generated such a complex model that we finally used only the 5% quality of the scanned point cloud for producing the mesh, as the full quality scan was not manageable on our workstations. This scanner had not captured the texture or color information on the object.

In our later work we used the hand held laser scanners, such as Z Scanner 900, for scanning the bronze portion of the monument to Franz Ferdinand and Sophie (Figure 3). In this case we used the full quality scanned data, while the post processing was done in the scanner portable software, as well as in 3ds max.

Figure 3. Laser scanning of the monument fragment

Laser scanning has not proved suitable for objects with hollow parts, as the scanners could not capture those parts with great accuracy.

IV. 3D MODEL CREATION

Although our first 3D model of a cultural heritage object, stećak from Donja Zgošća, was generated using laser scanning [1], most of the models in the next projects were created using classical modeling techniques, mainly because we had no remains of those objects available for data capture. We created models of objects that do not exist any more, as they were destroyed or demolished, such as Isa bey's endowment [2,3], the Church of the Holy Trinity in Mostar [4], Vizier's Konak in Travnik [5] (Figure 4).

Figure 4. a) The Church of the Holy Trinity in Mostar, above, b) Vizier's Residence in Travnik, below

The models of Roman objects created for enhancing the European museum exhibition "Keys to Rome", were also created using classic modeling techniques, but some preserved fragments were digitized using photogrammetry and added to the geometry of the models.

The 3D model of the monument to Franz and Sophie, (positioned until 1919 on the edge of the Latin bridge in Sarajevo), was created as a combination of laser scanning the preserved parts and modeling the rest of the object. The laser scanned data were subjected to some post processing in order to serve as a part of the 3D geometry [6].

3D models of medieval gravestones and stone monuments in the "Digital Catalog of Stećaks" project were created by photogrammetry and improved by post processing in 3ds max [7]. The post processing resulted in creating accurate geometry of the objects with all surface characteristics built in.

Virtual reconstruction of the Sultan Murat IV fortress in Sjenica, Serbia, was created only from the oral history [13]. This fortress was completely demolished in 20th century and today at its place stands an elementary school. We interviewed a person who remembers the fortress object from his childhood and created the 3D model based on the information from these stories.

V. APPLICATIONS

The main application of our virtual cultural heritage project was presentation of the valuable objects and sites to the general public through Internet. Our aim is to enable the visitor to travel to the past and learn the historical facts, while being immersed in the virtual environment in a way as close as possible to reality.

For this purpose we used several web 3D technologies, such as VRML, x3D and Unity, as well as some that only simulate 3D environments, such as Flash. One of the most important aspects of user immersion in the virtual environment is the quality of interaction that the environment provides while the user is browsing around. Apart from navigation, we introduced digital storytelling in our applications, in order to teach our visitors the historical context of presented cultural heritage objects and stories about events and characters related to them. Combination between navigation and storytelling was a

very interesting variable for evaluating the user's perception of the environment.

A. *Interactive Virtual Environments with Storytelling*

The first concept of interactive virtual environments with storytelling we implemented in the "Virtual Museum of BH Traditional Objects" [8]. This project is presenting the exhibition of objects from every day's life of people from different cultures that have been living in Bosnia and Herzegovina for centuries. These objects are not easy to understand for Internet visitors from all over the world if they are not accompanied by stories about their origin, purpose and way of use. Therefore, we presented every object through text, gallery of photos, digital story and an interactive 3D model. The objects are accessed by clicking on panels representing each of them in a joint virtual environment.

The main disadvantage of this concept was that users do not visit all available objects, but after some time they get tired and leave the virtual museum. For that reason, in our next virtual museum project, the Sarajevo Survival Tools (Figure 5), we introduced a digital story guiding the visitors through the exhibition [9].

Figure 5. Sarajevo Survival Tools virtual museum

In the Virtual Museum of Bosniak Institute, we experimented with reducing the freedom of movement to the visitors by implementing the virtual environment in form of rendered images with hotspots, while at the same time providing the audio stories about the objects and collections. It was interesting to find out that the users were not aware of movement limitations, as they were paying more attention to the stories [10].

The next step in our research of interactive digital storytelling for cultural heritage was the Isa bey's endowment project. In that project we implemented two concepts: interactive computer animation and spatial interactive storytelling [3]. In the recreated environment were the objects that the founder of Sarajevo, Isa bey Ishaković, has built for public benefit on the bank of river Miljacka. There was a soup kitchen, water mills, an accommodation facility and a derwish tekke[1]. These objects do not exist any more and the surroundings have been significantly changed. Within the virtual reconstruction we recreated the derwish ritual «zikr», that was going on inside the tekke. The user is placed in the middle of a computer animation of the ritual, where the narrator is explaining the characteristic parts, with a possibility to click on highlighted objects and see them explained in more detail in digital stories. Outside of the tekke, the user listens to the audio story about Isa Bey's endowment and, while navigating around the individual objects, the stories about them are triggered by proximity sensors.

Advantages and drawbacks of all mentioned interactive storytelling concepts led us to the work in progress: the recursive interactive story guided virtual museum [11]. The aim of this concept is to convey the information about cultural heritage to the users according to the amount of time they plan to dedicate to the visit. The stories are branching to more detailed stories using hyper-video.

B. *Augmented Reality*

Augmented reality (AR), according to Merriam-Webster Dictionary, is "an enhanced version of reality created by the use of technology to overlay digital information on an image of something being viewed through a device (such as a smart phone camera)". In virtual presentation of cultural heritage this technology has a great potential, providing the mobile device users the additional information over their camera images. In that way we can recreate appearance of our surroundings from different time periods and bring the forgotten cultural heritage objects back to collective memory.

Our Sarajevo Time Machine project was a pilot application created to show the potential of AR to cultural heritage professionals. We created simple 3D models of 6 selected objects in Sarajevo down town and stories about them. This content can be viewed on the mobile device display using the Layar open source AR platform. In this way the users can see on their screens the objects that used to exist in places where they stand and there is no physical trace of them today. The objects are shown in the 4D interactive map on the project web site, according to the time periods of their origin (Figure 6).

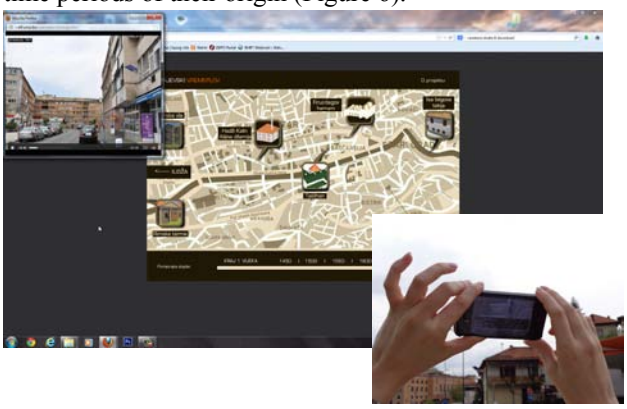


Figure 6. Sarajevo Time Machine project

VI. ENHANCING MUSEUM EXHIBITIONS

Our successful virtual museum projects earned us the invitation to join the EU FP7 Network of Excellence "Virtual Museum Transnational Network V-MusT.net". This network has a goal of establishing a new research discipline that would integrate the knowledge in application of digital technologies for enhancing museum exhibitions and presentation of cultural heritage. In the last year the consortium has organized an international

[1] Derwish are Islamic sufi believers who gather in places called „tekke" to discuss about religion and perform the ritual prayer called „zikr"

multimedia exhibition on Roman culture "Keys to Rome" [12]. This exhibition is held at the same time in Rome, Amsterdam, Alexandria and Sarajevo, with aim to present different parts of Roman Empire through a combination of physical exhibits and digital content. Roman objects that remained in all four locations were virtually reconstructed and the 3D models of museum artifacts were positioned inside. Using digital installations, the exhibition visitors could walk through Roman villas and temples and collect objects connected through a common digital story. Besides the exhibits from the physical location of the exhibition, they could see the other three exhibition locations virtually.

In Sarajevo the exhibition was held in the newly restored City Hall (Figure 7). It contained the physical exhibits from the archaeological sites in Ilidža near Sarajevo kept in the National Museum of Bosnia and Herzegovina and the Museum of Sarajevo. We recreated the Roman villa and thermal baths from Ilidža, as well as the Early Christian basilica from Cim near Mostar. The visitors could browse these virtual environments using the Admotum natural interaction setup (Figure 8). When they would find an object of interest they could transfer it to the Holobox installation and explore it in detail using the Leap motion sensor. All physical exhibits had their virtual models connected with exhibits from other three locations within the Matrix mobile application.

Figure 7. Keys to Rome exhibition, location Sarajevo

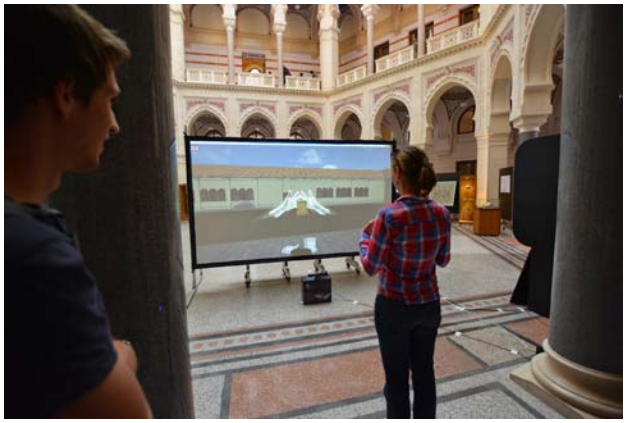


Figure 8. Admotum digital installation

VII. User Evaluation

One of the most important methods for measuring the success and quality of virtual cultural heritage projects is the user evaluation. We are using the customer satisfaction methodology with questionnaires and the qualitative analysis of user interviews.

Ritual Zikr in Isa Bey's tekkeh

interactive animated story

Evaluation - film + interactive story

Personal data
1. How old are you?
2. Do you have problems with your eyesight?
3. Do you have a hearing problem?
4. Whether you are an experienced computer user? For what you use computer the most (work, games, mail / Skype, etc.)?

After watching the film, and interactive stories users need to respond (preferably in detail) the following questions:

1. Have you heard of Isa Bey tekke?
2. Do you know what a tekke?
3. Which order of dervishes gathered in this tekke?
4. What did you learn from the examined films / animations?
5. Did you feel like you were in a tekke?
6. What is missing?
7. What bothered you?
8. What did you like the best?

9. Have you had problems with navigation in the interactive environment?
10. Have you used a site map (house in the upper left corner of the screen)?
11. Would you be back again to look at what you missed?

12. What is better for learning about the object? A film or interactive form?
13. Why?
14. Do you have any recommendations to improve the interactive form?

Thanks for your cooperation!

Figure 9. Example of a questionnaire for user evaluation

The questionnaires we use usually contain questions related with the following topics: general information on the user (age, profession, familiarity with use of computers, sight and hearing impairments, familiarity with the reconstructed object), usability of the application (UI, navigation, immersion, interaction), obtained information (quality of digital storytelling) and positive/negative impressions (Figure 9).

Interviews were conducted in such a way that the users would talk about their experience with the application, being occasionally directed with some questions from the interviewer about the same topics covered in the questionnaires. The qualitative data obtained in the interviews were converted into quantitative form by the process of data coding. Coding extracts values for quantitative variables from qualitative data (interviews and questionnaires) to perform quantitative or statistical analysis [14] (Table 1). The process of coding does not affect data subjectivity or objectivity. Based on the research question we need to answer through the particular user study, we formed a number of hypotheses and explored if they were confirmed by the coded answers. Such a user evaluation of the interactive "zikr" ritual animation in the Isa Bey's endowment project is presented in [15].

Based on all our user studies, we can conclude that the feedback of the users to our virtual cultural heritage projects is very positive. They appreciate finally being able to visualize the remains of objects they visit at the archeological sites, to virtually explore those objects and learn about their history, about events and characters related to them. The users familiar with computer games

easily adjusted to these applications, while the less experienced users expressed some concerns about navigation methods that we have taken in account and corrected in the next projects. The use of digital storytelling has significantly enhanced the feeling of immersion and the amount of obtained information on the objects.

Question	Code	Possible value	Answer
Have you heard of Isa Bey tekke?	I1	yes no	15 5
Do you know what is a tekke?	I2	yes no	18 2
Which order of dervishes gathered in this tekke?	I3	right wrong	16 4
Did you feel like you were in a tekke?	E1	yes no	15 5

Table 1: Category 1 questions

Question	Code	Possible value	Answer
Have you had problems with navigation in the interactive environment?	N1	yes no	4 11
Have you used a site map (house in the upper left corner of the screen)?	N2	yes no	12 3
Would you be back again to look at what you missed?	N3	yes no	14 1
What is better for learning about the object? A movie or interactive form?	E2	movie inter. form equal	2 6 2

Table 1. Example of data coding in qualitative user evaluation

The user studies showed us what aspects of virtual environments were most important for their perception and it was not always the same as what we considered as important in the process of project development. For example, they considered more important the quality of navigation then the realism of displayed environments.

VIII. CONCLUSION

The overview of virtual cultural heritage projects implemented by Sarajevo Graphics Group shows our experience in using different digital technologies for cultural heritage digitization, presentation and preservation. We described the main methods of data capture, 3D model creation and presentation of those models in various web 3D and augmented reality technologies.

Considering the rapid development of 3D technologies, there is a serious issue of technologies becoming obsolete. Sustainability and life span of virtual cultural heritage is a problem addressed by many scientists and research groups. There is still no universal solution except structuring the projects in such way that the technology related parts can easily be upgraded.

A very positive feedback of our users shows that the virtual presentation of cultural heritage is appreciated and offers the knowledge on our past in a modern and up-to-date way. Through our projects we brought many forgotten objects back to the collective memory. Virtual travel to the past was equally exciting for all generations and backgrounds of our users.

There is still a lot of work to do in discovering the most immersive way of cultural heritage presentation. We find the interactive digital storytelling as a powerful tool for conveying information to the virtual and real visitors. The use of digital technologies for enhancing the museum collections results with development of a new profession – digital curators. They will have enough knowledge about the potential of digital technologies to invent new ways of organizing museum exhibitions. Virtual museums will not replace the physical museums, but attract more visitors to them. Virtual cultural heritage projects will enhance the development of cultural tourism. Better understanding of the past will improve our perception of the present.

ACKNOWLEDGMENT

I would like to express my gratitude to my former assistants Aida Sadžak, Belma Ramić Brkić and Vedad Hulusić for their relentless efforts in the difficult years when we were establishing our Group. I thank also my former undergraduate and masters students Anis Zuko, Vanja Jovišić, Sanda Šljivo, Goran Radošević, Merisa Huseinović, Izabela Skalonjić, Mohamed El Zayat and Bojan Kerouš for their great work and creativity in recreation of our cultural heritage. Thanks also to our collaborators from museums and cultural heritage institutions for providing us help and expertise in the fields unfamiliar to us. Finally, I am eternally grateful to Professors Alan Chalmers and Roberto Scopigno for recognizing the quality of our work and introducing us to the international scientific community.

REFERENCES

[1] Selma Rizvic, Aida Sadzak , Emir Buza, Alan Chalmers - Virtual reconstruction and digitalization of cultural heritage sites in Bosnia and Herzegovina, Cetinje, SEEDI 2007, *Review of the National Center for Digitization*, 82 – 90, Publisher: Faculty of Mathematics, Belgrade, Serbia,ISSN: 1820-0109J

[2] Selma Rizvić, Aida Sadžak, Anis Zuko, Isa bey's Tekija in Sarajevo - reviving the reminiscence of the past, *Review of the National Center for Digitization*, Publisher: Faculty of Mathematics, Belgrade, Issue: 15/2009, pg 64-72, ISSN: 1820-0109

[3] S. Rizvic, A. Sadžak, M. El Zayat, B. Žalik, B. Rupnik, N. Lukač, Interactive Storytelling About Isa Bey's Endowment, *Proceedings of SEEDI* 2013, Zagreb, Croatia

[4] B. Ramic-Brkic, Z. Karkin, A. Sadzak, D. Selimovic & S. Rizvic, Augmented Real-Time Virtual Environment of the Church of the Holy Trinity in Mostar, *Proceedings of VAST* 2009, ISBN 978-3-905674-18-7, pg 141-148

[5] V. Jovišić, Virtual Reconstruction Of The Viziers' Konak In Travnik, Master Thesis, Sarajevo School of Science and Technology, 2009

[6] G. Radošević, S. Rizvić, Spomenik Ferdinandu i Sofiji - od laserskog skena do interaktivnog 3D modela, *Drugi medjunarodni simpozij "Digitalizacija kulturne bastine Bosne i Hercegovine, Sarajevo"*, maj 2010.

[7] S. Rizvić, A. Sadžak, Digitalni katalog stecaka, *Drugi medjunarodni simpozij "Digitalizacija kulturne bastine Bosne i Hercegovine"*, Sarajevo, maj 2010

[8] S. Rizvić, A. Sadžak, Multimedia techniques in virtual museum applications in Bosnia and Herzegovina, *International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2011, pp 1-4 ISBN 978-1-4577-0074-3

[9] S. Rizvic, A. Sadzak, V. Hulusic, A. Karahasanovic, Interactive digital storytelling in the Sarajevo survival tools virtual environment, *Proceedings of the 28th Spring Conference on Computer Graphics*, Pages 109-116, ACM New York, NY, USA ©2012, ISBN: 978-1-4503-1977-5

[10] S. Šljivo. Audio Guided Virtual Museums, in Proceedings of Central European seminar on Computer Graphics, Smolenice, Slovakia, 2012

[11] S. Rizvic, Story Guided Virtual Cultural Heritage Applications, *Journal of Interactive Humanities*, Vol. 2: Iss. 1, Article 2, ISSN: 2165-7564, 2014

[12] International multimedia exhibition on Roman culture, Keys to Rome, www.keys2rome.eu

[13] I. Skalonjic, Interaktivna miksmedijalna virtuelna rekonstrukcija tvrđave Sultana Murata IV u Sjenici, *Masters thesis, Faculty of Electrical Engineering Sarajevo*, 2014

[14] C.B.Seaman. Virtual Qualitative Methods in Empirical Studies of Software Engineering. IEEE Transactions on Software Engineering, Vol. 25, No. 4, 1999.

[15] M. Huseinovic, R. Turcinhodzic, Interactive Animated Storytelling In Presenting Intangible Cultural Heritage, *In Proceedings of the Central European Seminar on Computer Graphics*, Smolenice, Slovakia, 2013

Comparative Analysis of Local and Global Innovation of Knowledge Sources in Standardized Subfields of Health Care Technology

Živadin Micić*, Marija Blagojević*
* Faculty of Technical Science Čačak - University of Kragujevac/ Department of IT, Čačak, Serbia
e-mail: micic@kg.ac.rs; marija.blagojevic@ftn.kg.ac.rs

***Abstract* — This paper presents a comparative analysis of knowledge sources innovation on the examples of standardized subfields of health care technology. Based on standardized quantity, value and intensity of innovation of knowledge sources, the clusters, as well as the frequency of innovation in analyzed standardized subfields, have been defined. What follows is the realization of a *knowledge base system* (KBS) in practice, with the insurance of resources for the quality of final products. This is enabled by using PDCA methodology, on the (standardization) platform, with classified areas of technology. On the examples of health care technology, with11 standardized subfields, original trend lines (towards the KBS) have been formed. The study results enable further development and application of the methodology, as well as standardized solutions to the issues of health care.**

I. INTRODUCTION

The paper deals with a comparative analysis of international (ISO/IEC) and local (SRPS – label for standards in Serbia, [1]) knowledge sources in the subfields of health care technology. Knowledge pathways in this field differ from those in other standardized technologies. One of the objectives of the work in the field of health care technology is to determine the importance of local in relation to global knowledge sources. According to the international classification of standards (ICS1 - classification of the first level), health care technology is classified in 11 subfields of the second level (ICS2 = 11.xy0, [2]):

11.020 Medical sciences and health care facilities in general,

11.040 Medical equipment,

11.060 Dentistry,

11.080 Sterilization and disinfection,

11.100 Laboratory medicine,

11.120 Pharmaceutics,

11.140 Hospital equipment,

11.160 First aid,

11.180 Aids for disabled or handicapped persons,

11.200 Birth control. Mechanical contraceptives and

11.220 Veterinary medicine.

According to [3], "Quality is free" philosophy was present in the previous century. Nowadays, standards are extremely costly.

A. The initial hypotheses and objectives of the research

Initial hypotheses are proven and the research objectives are realized in the standardized subfields of health care technology through the PDCA quality loop:

Hypothesis_1 - P (Plan) Planning and prediction of necessary future resources and financial requirements are possible for the evaluated units of knowledge sources and responsibilities for each individual subfield (for ICS2 = 11.xy0) and for all of them together,, in subcommittees and development stages of new projects (from a practical point of view – ISO/IEC and SRPS),

Hypothesis_2 - D (Do) Research and evaluation of knowledge sources units enable the formation of explicit mathematical relations, as well as regression trend lines of knowledge (from a practical point of view),

Hypothesis_3 - C (Check) Definition of clear correlations between obligations and knowledge with the intensity of innovation of valued knowledge sources units is possible on the relations between ISO and SRPS (clustering),

Hypothesis_4 - A (Act) It is possible to define the relations between the continuous (according to the PDCA) and discontinuous knowledge innovation of individuals, with the ultimate goal of improving the teamwork (knowledge base system) and innovation of industrial products on the platform of SRPS and ISO standardization.

II. RESEARCH METHODOLOGY AND FRAMEWORK

The statistical methodology of dynamic analyses and deductive - inductive reasoning methods were used for predicting the future development and innovation of the pragmatic framework. Methodologically, statistical indices were formed for the comparison of ISO - SRPS relations in the field of health care technology (ICS1 = 11) with other fields of human endeavor, including: Quantity indices (Iq), value index (Iv) and index of quantitative variation for ranking (Iqi).

The PDCA methodology, statistical research and multicriterium analyses have been applied.

A. Study framework

Quantity indices (Iq), defined and determined for both ISO and SRPS, refer to: Samples (Iqs), Published standards (Iqp), Standards Under development (Iqu), Standards Withdrawn from use (Iqw), Deleted projects

(Iqd), Innovations in various stages of development (Iqi = $Iqu_{/2012}$) - for the full previous calendar year. In general, for the population Iqs equation (1) is derived:

Iqs = Iqp + Iqw + Iqd + Iqu (1)

For the complete "game"-research (browsing, analysis, systematization and presentation of the results) two original JAVA applications [4] were used for research purposes, as the examples of IT innovations of software products in comparable samples in the subfields of health care technology (za ICS1 = 11), of two application levels, ISO and SRPS:

1) the first JAVA application browses (analyzes) ISO standardized unit base ($Iqs_{/ISO/2014.01}$ ≈ 42000 collective-global innovations on the examples of standards, [2]) and sample of $Iqs_{/11/ISO/2012.12}$ = 1849 units out of total $Iqs_{/ICS=1-99/ISO}$,

$Iqs_{/11/ISO/\mathbf{2015.01}} = Iqs_{/11/ISO/P5} =$ **2054** (1.1)

where the total quantity of the global (ISO) sample is $Iqs_{/1-99/ISO/2013.01} \geq 41141$.

2) the second IT application gathers (and analyzes) data on local SRPS standardization ($Iqs_{/SRPS/2014.01}$ ≈ 34000 local "innovation", [1]) and sample of $Iqs_{/11/SRPS/2012.12}$ = 882 knowledge units out of $Iqs_{/ICS=1-99/SRPS}$.

$Iqs_{/11/SRPS/\mathbf{2015.01}} = Iqs_{/11/SRPS/P5} =$ **1089** (1.2)

where the total quantity of the local (SRPS) sample is $Iqs_{/1-99/SRPS/2013.01} \geq 31199$.

B. Trend lines in PDCA quality loop

All the results obtained so far (2012/01 → $P_2D_2C_2A_2$, 2013/01 → $P_3D_3C_3A_3$ - - - 2014/01 → $P_4D_4C_4A_4$ and 2015/01 → $P_5D_5C_5A_5$) have been evaluated by the analysis, for the explored subfields and the field taken as a whole: $Iv_{/ISO}$, $Iv_{/SRPS}$. The results are presented aggregately and by trend lines using graphical displays including:

a) time aspects for the whole research period – according to the year of publication, ∑Iv/year, as well as

b) regression trend lines (Exponential, Linear, Logarithmic and Polynomial), according to the data from the previous years of the XXI century (for example, 2008-2012) and defined regression equations $y_{i/SRPS}$, Ivi/year.

$\mathbf{y_{11/ISO/2000-2011/P2} = 222.8\ x + 2823}$ (2.1)

$\mathbf{y_{11/SRPS/2000-2011/P2} = 202\ x}$ (2.2)

C. Definition and checking of innovation intensity

The results of the aforementioned analyses enable the creation of an original methodology for the comparison of innovations in all technologies. This is realized by defining the indices as criteria for the clustering of appropriate subfields and/or fields.

Indices of time intensity of innovation (Iti) enable clustering by subfields and continuous periodic updating of knowledge. The periodicity of innovation index (Iti) is defined on the basis of quantitative indices Iqi (table I - column 5, column 8 and table II - column 4), which is in the immediate multi-criteria qualitative and financial dependence (3).

$Iqi = Iqu_{/ISO/year+1} + Iqp_{/SRPS/year}$ (3)

Iti = 0, for Iqi = 0 (3.1)

Iti = 1, for 1 ≤ Iqi < 12 (3.2)

Iti = 2, for 12 ≤ Iqi ≤ 50 (3.3)

Iti = 3, for 50 < Iqi ≤ 250 (3.4)

Iti = 4, for 250 < Iqi (3.5)

The values of periodic checks (Check) of the research for practice (table II - column 3 for the Iti and column 4 for the Iqi): Iti = 0 — annual Check, relation (3.1), Iti = 1 — annual-yearly Check, relation (3.2), Iti = 2 — monthly Check, Iti = 3 — weekly Check or Iti = 4 — daily Checks, relation (3.5), are assigned to this index [4], [10].

Methodologically, statistical indices have been formed for the comparison of ISO/IEC - SRPS relations in this field (ICS1 = 11) with other fields.

D. Comparison of intensity inovativnsoti and innovation of KBS

The results obtained so far have been evaluated by a multi-criteria analysis at the end of the previous or in the beginning of next year (2012/12, table I and table II). Quantity and value indices enable internal comparisons in researched subfields and in the field as a whole, and external comparisons with other fields. The methodology enables the research and comparison of the results with other fields - as according to [4].

After the formation of the knowledge base (KB), the next stage in the PDCA spiral is the formation of a knowledge base system (KBS). The result is a timely updating of the *Knowledge Base System - KBS*, in the adequate field/subfield (KBSti/ics ≈ function (Iqi and Iti) including PDCA at the time moment (t)) according to the relation (4) for the excellence model.

KBSti/ics ≈ ∑(Iqp + fun.(Iqi, Iti) & PtDtCtAt)/ics (4)

III. RESULTS

The results of the analysis of standardized knowledge sources are presented by comparative indices of quantity, value and intensity of innovation for each of 11 subfields and health care technology field in general. Statistical samples (Iqs) with comparative results, other indices of quantity and value should further "explain" knowledge pathways in health care technology, table I.

Table I

Quantitative indices (Iq), parallel ISO – SRPS, for ICS1 = 11, 2012/December

I	Sub/ field	Samples		**Developed (Iqi)**				Withdrawn	
		Iqs		**Iqu**		**Iqp/2012**		**Iqw**	
	ICS	**ISO**	SRPS	**ISO**	SRPS	**ISO**	SRPS	**ISO**	SRPS
1)	2)	3)	4)	5)	6)	7)	8)	9)	10)
1	**11.020**	**7**	**5**	**3**	**0**	**0**	**0**	**3**	**2**
2	**11.040**	**1047**	**442**	**125**	**5**	**53**	**137**	**432**	**30**
3	**11.060**	**369**	**171**	**34**	**7**	**11**	**15**	**178**	**23**
4	**11.080**	**96**	**107**	**17**	**1**	**2**	**3**	**34**	**28**
5	**11.100**	**105**	**63**	**8**	**0**	**5**	**4**	**33**	**12**
6	**11.120**	**10**	**9**	**0**	**0**	**0**	**3**	**4**	**2**
7	**11.140**	**16**	**13**	**0**	**0**	**1**	**4**	**5**	**1**
8	**11.160**	**0**	**18**	**0**		**0**	**5**	**0**	**7**
9	**11.180**	**146**	**44**	**15**	**0**	**10**	**15**	**38**	**3**
10	**11.200**	**53**	**9**	**1**	**0**	**1**	**0**	**38**	**0**
11	**11.220**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**
XII	**∑ 11**	**1849**	**882**	**203**	**13**	**83**	**186**	**765**	**108**

The results of the analysis of knowledge standardization are presented in Table I, Table II and the

accompanying charts. Graphical displays are given only for the subfields that belong to the second group and clusters with higher intensity of innovation (according to table I). Graphical displays include two parts:

a) the analyses of all current standards (Std), corrections (Cor), amendments (Amd) and projects under development,

b) the analyses, choice and displays of some of the regression trend lines , as well as a link with the columns (2 to 8), table II.

In order to provide clearer graphical displays of the results, only analyses of the subfields of the second-level classification (or ISC2, table I and table II) are presented in detail.

The displays of the details of the analysis in the subfields of the third level of classification have been excluded (i.e. there are no more details for the ICS3).

Table II

Innovation index (Iti), indices of value (Iv), parallel ISO - SRPS, for the ICS1 = 11, 2012/December

	Subfield			∑Iv/2012.12 (CHF)		Ivp/2012 (CHF)	
I	ICS	Iti	Iqi	∑Iv/iso	∑Iv/srps	Ivp/iso	Ivp/srps
1)	2)	3)	4)	5)	6)	7)	8)
1	**11.020**	**1**	**3**	**352**	**193.54**	**108**	**0**
2	**11.040**	**4**	**262**	**39978**	**12868.23**	**5112**	**3516.36**
3	**11.060**	**3**	**49**	**10590**	**3560.49**	**824**	**312.47**
4	**11.080**	**2**	**20**	**4876**	**3019.59**	**270**	**93.45**
5	**11.100**	**2**	**12**	**5840**	**1643.92**	**578**	**88.37**
6	**11.120**	**1**	**3**	**292**	**192.02**	**0**	**52.41**
7	**11.140**	**1**	**4**	**908**	**296.97**	**16**	**104.20**
8	**11.160**	**1**	**5**	**0**	**432.88**	**0**	**110.69**
9	**11.180**	**2**	**30**	**8854**	**1115.50**	**1318**	**399.59**
10	**11.200**	**1**	**1**	**1040**	**126.49**	**16**	**0**
11	**11.220**	**0**	**0**	**0**	**28.15**	**0**	**0**
XII	**∑ 11**	**4**	**389**	**72730**	**23477.78**	**8242**	**4677.54**

A. *Daily intensity of innovativeness*

Out of the mentioned 11 subfields (ICS2 = 11.x, table II) the subfield – Medical equipment **(**ICS2 = 11.040) belongs to a cluster of **daily** intensity innovativeness.

Medical equipment subfield belongs to the fourth group (table I – column 3). There is the highest number of documents in all forms in this subfield: samples (Iqs), under development - drafts (Iqu), withdrawn (Iqw), and deleted (Iqd). The analysis results are presented graphically in Figure 1:

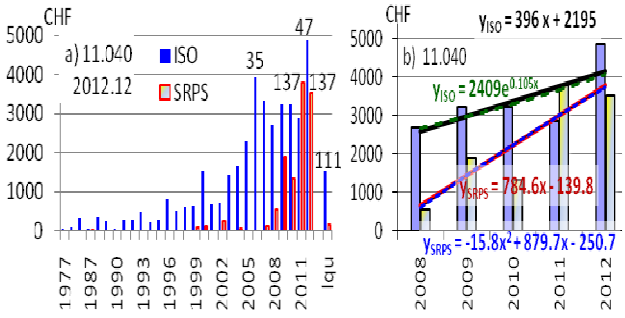


Figure 1. Analysis of the results for ICS2 = 11.040 – Medical equipment

a) with all available samples from the period from 1977 to 2012, Figure 1a,

b) with the trend of planned future needs, according to the relations (5.1) and (5.2), Figure 1b.

$$y_{11.040/ISO/2008\text{-}2012} = 396\,x + 2195 \quad (5.1)$$

$$y_{11.040/SRPS/2008\text{-}2012} = 784.6\,x - 139.8 \quad (5.2)$$

An example of the analysis of this subfield labeled as ICS2, is carried out in depth through the analysis of the next level - ICS3. Available samples for the analyses at the level of this subfield include $Iqs_{/11.040/2012}$ = 1489 units (ISO - 1047 and SRPS - 442), distributed by the subfields classified according to ICS3:

11.040.01 Medical equipment in general

11.040.10 Anesthetic, respiratory and reanimation equipment **Including medical gas installations*

11.040.20 Transfusion, infusion and injection equipment **Including blood packs *Syringes, needles and catheters, see 11.040.25*

11.040.25 Syringes, needles and catheters

11.040.30 Surgical instruments and materials **Including surgical dressings, sutures, etc.*

11.040.40 Implants for surgery, prosthetics and orthotics **Including pacemakers *Ophthalmic implants, see 11.040.70*

11.040.50 Radiographic equipment **Including radiographic diagnostic and therapy equipment *Dental, medical and industrial radiographic films, see 37.040.25*

11.040.55 Diagnostic equipment **Including medical monitoring equipment, medical thermometers and related materials*

11.040.60 Therapy equipment

11.040.70 Ophthalmic equipment **Including ophthalmic implants, glasses, contact lenses and their cleaning products*

11.040.99 Other medical equipment.

B. *Weekly intensity of innovativeness*

Out of the mentioned 11 subfields, the subfield ICS2 = 11.060 (table II) belongs to a cluster of **weekly** intensity innovativeness. The examples of the results of the analysis are presented for the subfield – Dentistry.

Dentistry subfield belongs to the third group (the only subfield with weekly intensity of innovation). Figure 2 shows the results of the analysis of knowledge sources ISO - SRPS (comparatively) in this sub-field for ICS2 = 11.060:

a) with valid published standards for use, from 1982 to 2012, Figure 2a,

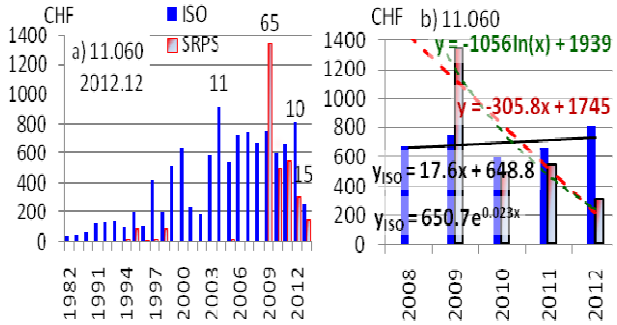


Figure 2. Analysis of the results for ICS2 = 11.060 – Dentistry

b) with trend lines of planned (annual) future needs, i.e., with very close regression lines, Figure 2b, the linear

relation (6.1) is close to the exponential - "global", and "local" - logarithm is close to the relation (6.2).

$$y_{11.060/ISO/2008\text{-}2012.12} = 17.6x + 648.8 \quad (6.1)$$

$$y_{11.060/SRSPS/2008\text{-}2012.12} = -305.8x + 1745 \quad (6.2)$$

An example of this subfield analysis, in depth goes through the analysis of the next level - ICS3:

11.060.01 Dentistry in general

11.060.10 Dental materials

11.060.15 Dental implants **Including dentures*

11.060.20 Dental equipment **Dental radiographic films, see 37.040.25 *Toothbrushes and dental floss, see 97.170*

11.060.25 Dental instruments.

C. Subfields of monthly intensity of innovativeness

Out of the mentioned 11 subfields (ICS1 = 11) the cluster of monthly intensity innovativeness includes three subfields: 11.080, 11.100 and 11.180 (table II). What follows is a presentation of the analysis of the results in the three above mentioned subfields.

C.1 Subfield – Sterilization and disinfection (11.080)

The results of the analysis of standardization in Sterilization and disinfection subfield are presented in Figure 3:

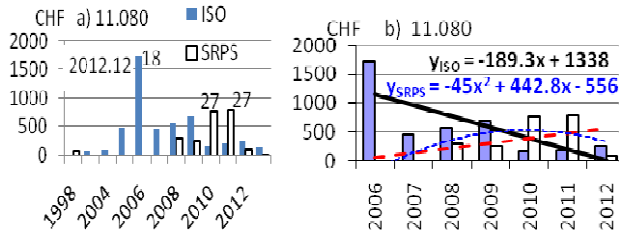


Figure 3. Analysis of the results for ICS2 = 11.080 – Sterilization and disinfection

a) all analyzed samples were created in the period from 1991 to 2012,

b) with the trend of planned needs (Figure 3b): for the ISO according to (7.1), for the SRPS according to (7.2), since the linear relation $y_{11.080/SRPS/2006\text{-}2012.12} = 82.55x - 15.53$ is not realistic.

$$y_{11.080/ISO/2003\text{-}2012.12} = -12.72x + 541.2 \quad (7.1)$$

$$y_{11.080/SRPS/2006\text{-}2012.12} = -45.03x^2 + 442.8x - 555.9 \quad (7.2)$$

C.2 Subfield – Laboratory medicine (11.100)

The results of the analysis in Laboratory medicine subfield are presented in Figure 4:

a) without currently valid standards before 1985, while most of the samples were formed in the XXI century (2002-2012, and SRPS 2007-2012),

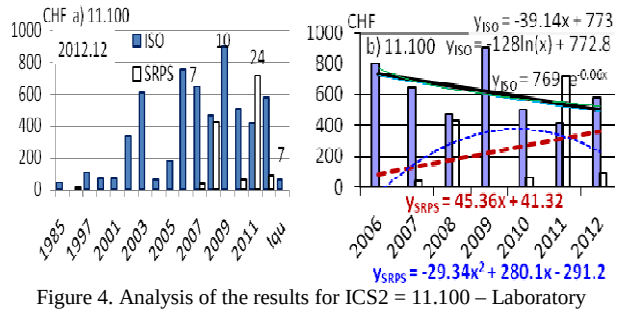


Figure 4. Analysis of the results for ICS2 = 11.100 – Laboratory medicine

b) with all regression equations of the global trend and ISO development close to the relation (8.1), but with a debatable local trend(SRPS): with a linear regression line ($y_{SRPS} = 45.36x + 4.32$) of theoretical character, since the number of projects under development $Iqu_{/SRPS} = 0$, with the trend of planned (annual) future needs, according to the relation (8.2), Figure 4b.

$$y_{11.100/ISO/2006\text{-}2012.12} = -39.14x + 773 \quad (8.1)$$

$$y_{11.100/SRPS/2006\text{-}2012.12} = -29.34x^2 + 280.1x - 291.2 \quad (8.2)$$

C.3 Subfield – Aids for disabled or handicapped persons (11.180)

Aids for disabled or handicapped persons subfield belongs to the second group (as well as subfields 11.080 and 11.100), including the assistance to the elderly. The results of the analysis are presented in Figure 5:

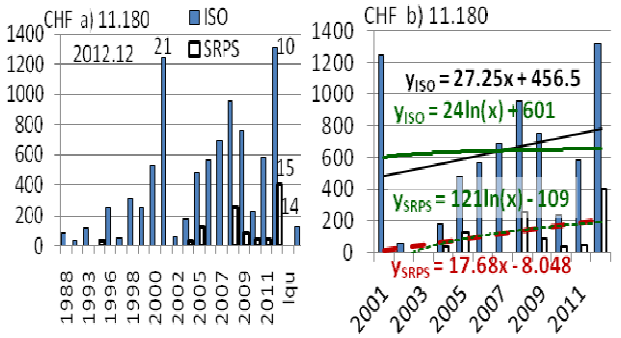


Figure 5. Analysis of the results for ICS2 = 11.180 – Aids for disabled or handicapped persons

a) without currently valid standards before 1988, with the majority of the samples created in the XXI century (2001-2012, and SRPS 2004-2012), Figure 5a,

b) with regression equations of the global trend and ISO development close to the relation (8.1), as well as with local trend (SRPS): with logarithmic regression line ($y_{SRPS} = 121\ln(x) - 109$), which is very close to the trend of planned(annual) future needs, according to the linear relation (9.2), Figure 5b.

$$y_{11.180/ISO/2001\text{-}2012.12} = 24\ln(x) + 601 \quad (9.1)$$

$$y_{11.180/SRPS/2001\text{-}2012.12} = 17.68x - 8.048 \quad (9.2)$$

D. Subfields of low intensity of innovation

Most of the mentioned 11 subfields (ICS2 = 11, table II) belong to the clusters of low intensity of innovation (annual level of innovativeness: 11.020, 11.120, 11.140, 11.160 and 11.200 and zero level of innovativeness 11.220).

Birth control. Mechanical contraceptives subfield belongs to the first group (as well as the aforementioned subfields). the amount of $Iqs_{/11.200/SRPS} = 9$ units in this subfield which is insufficient for statistical analysis. Other results and quantity of the ISO samples complete the analysis:

a) the results are presented graphically, aggregately, with all available samples from the period from 1995 to 2012, Figure 6a,

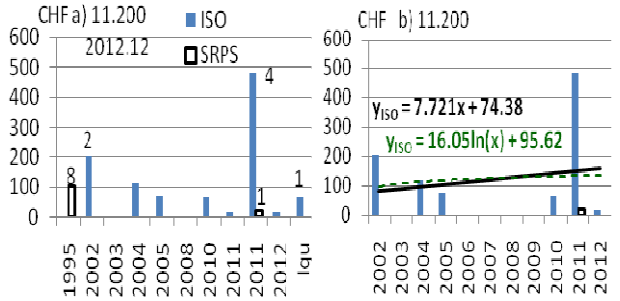


Figure 6. Analysis of the results for ICS2 = 11.200 – Birth control. Mechanical contraceptives

b) with linear regression line (y_{ISO} = 7.721x + 74.38), which is very close to the trend of planned future needs (based on global ISO development), according to the logarithmic relation (10.1), Figure 6b.

$$\mathbf{y_{11.200/ISO/2002\text{-}2012} = 16.05\ln(x) + 95.62} \quad (10.1)$$

IV. DISCUSSION OF THE RESULTS

The results show significant detail analyzed standardized units of knowledge sources, by subfields health care technology. In Figure 7a presents the typical results (Iv/year, Iqu) in the period from 1977 to 2011, for the analyzed area as a whole (ICS1 = 11). The obvious is negligibly small percentage of valid SRPS units created before 2008 - in the domain of mathematical statistical errors.

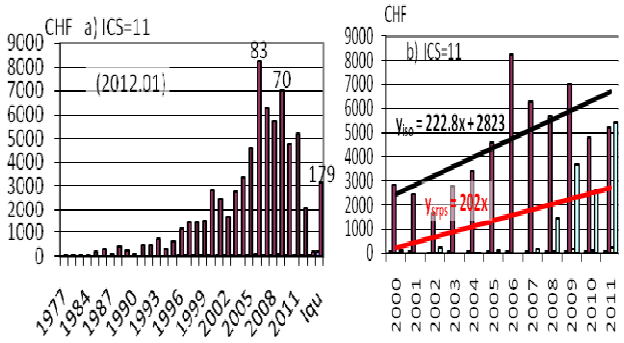


Figure 7. Examples of necessary resources and knowledge of all SRPS and ISO standards (ICS1 = 11)

$$\mathbf{y_{11/ISO/2000\text{-}2011} = 222.8\ x + 2823} \quad (11.1)$$

$$\mathbf{y_{11/SRPS/2000\text{-}2011} = 202\ x} \quad (11.2)$$

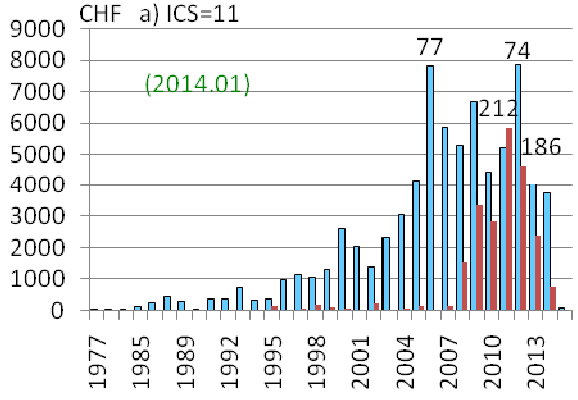


Figure 8 a) Examples of necessary resources and knowledge of all SRPS and ISO standards (ICS1 = 11) 2014.01

In 2012, new $Iqp_{/11/SRPS/2012.12}$ = 186 SRPS standards – which means that an innovation occurred each day! On the other hand, there is also a significant number of ISO innovations in 2012, as well as projects under development $Iqu_{/11/ISO/2012.12}$ = 203.

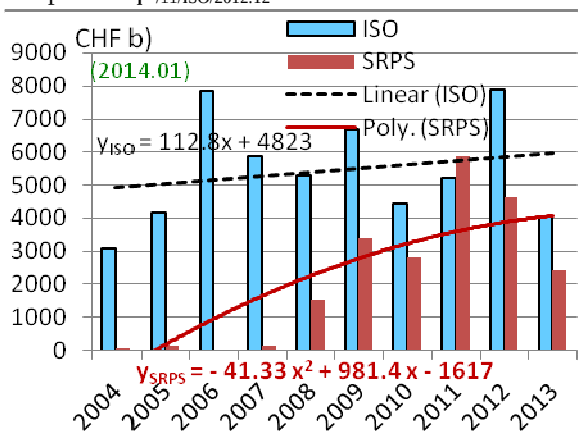


Figure 8 b) Examples of trend line for SRPS and ISO (ICS1 = 11) 2014.01

Based on these results, the evidence presented the initial hypotheses in PDCA: Plan – Do – Check – Act.

Plan

Future resources and financial needs can be presented for each subfield individually, and for the entire field. This is allowed by the established mathematical relations, presented trend lines (from a theoretical point of view), defined indices of value Iv/year, but also knowledge of the stages of new projects development (from a practical point of view - with Iqu).

Do

Compared with other fields, such as IT (ICS1 = 35), or IT industry subfield (ICS2 = 25.040), according to [4] and [10], the same relationship between valued standardized innovations is present in this field: there are less local knowledge innovations (SRPS) compared to the global (ISO).

Check

Innovation index (according to the presented results, Iti = 1.86 + 2.03) classifies this field in the cluster of daily intensity innovation. More specifically, in 2012 there were new $Iqp_{/11/SRPS/2012.12}$ = 186 SRPS standards. On the other hand, there was a significant number of ISO innovations in 2012, as well as projects under development $Iqu_{/11/ISO/2012.12}$ = 203 - there was an innovation each day!

Act

Presented methodology allows the comparison of the results with the results in other fields of work and standardized human activities.

The results in the above mentioned fields and subfields are comparable with the results of other published research. For example: in the field of Manufacturing engineering (ICS1 = 25) according to [5], in Health protection and safety (ICS1 = 13), according to [6], in Electrical engineering (ICS1 = 29) according to [7], in the field of IT (ICS1 = 35), according to [4], in IT applications subfield (ICS2 = 35.240), according to [8]), or for comparison with the already published results in the subfields such as Metallurgy (ICS1 = 77) according to [9], or knowledge innovation trends in the field of Railway

Engineering (ICS1 = 45) and Civil Engineering (ICS1 = 93) according to [10].

Compared with other papers dealing with the issue of existence of a standard framework for creation of standards [11], this paper deals with the possibilities of access to the standards, with a goal of their implementation for innovating the knowledge in PDCA - [12].

Unlike the papers [13], which present a comparative analysis of two advanced ICT nations, this paper gives the analyses of the relations of the standards of one nation (local SRPS) comparatively with the global ISO.

Activities on improving this paper, starting from the knowledge sources on 1.1.2014 and 1.1.2015,, and modeling the knowledge base towards the *knowledge base system*, will be presented in future paper.

V. CONCLUSIONS

Based on the multi-criteria analysis of the knowledge sources base and presented results on the examples of standardized subfields of health care technology (ICS1 = 11) and with the help of confirmed initial hypotheses, the conditions have been made for new results/papers at every stage of the PDCA methodology.

Plan-phase — Monitoring of the developmental phases of new projects is an important aspect of alternative solutions to this "issue", when planning and providing resources for the quality of services in *health care technology*. Proof of hypothesis 1.

Do-phase — Defined indices of quantity and value have enabled practical and standardized grouping by subfields, with better organization of duties and tasks, as well as comparisons with all other fields and subfields. Clustering in *health care technology* is the starting point for future papers in several directions, in relation to new study programs, formation of *knowledge base system*, etc.

The analysis of knowledge sources base in the field of *health care technology* has shown the relations between the local (SRPS) and international (ISO) knowledge. High values of defined indices indicate questionable opportunities for accessing individuals (i.e. difficult access to the sources of SRPS in Serbia).

Clear correlations between obligations and knowledge with the annual trends of innovating standardized sources of knowledge have been defined on the relations between ISO - SRPS by analyzed fields and subfields for the field of health care technology: $Ivp_{/11/ISO/2012}$ = 8242 CHF per year, according to Table II (or $Iv_{/11/ISO/2013}$ = ≈ 8000 CHF, according to relation 11.1) - continuously for ISO data and knowledge base units (which is comparable to $Ivp_{/11/SRPS/2012}$ = 4678 CHF - according to the trends of SRPS standards in Figure 7b, or relation 11.2, where: $Iqs_{/11/SRPS}$ = 882, $\sum Iqp_{/11/SRPS/2012.12}$ = 761, $Iqp/_{11/SRPS/2012}$ = 186, etc. An approximate ratio of all valued standardized sources of knowledge, SRPS - ISO, has been defined, or numerically $\sum Iv_{/11/SRPS/2012}$ = 23478 CHF compared to$\sum Iv_{/11/ISO/2012}$ = 72730 CHF, in early 2013.

Check-phase — Clearly defined indices of quantity, frequency and intensity of innovation in each subfield enable clustering with time periods of the implementation of the PDCA cycle quality loop (table II). This is a confirmation of hypothesis 3.

Act-phase — Relationship between the indices of value of continuous ($Ivp_{/11/PDCA}$) and cumulative knowledge innovation ($\sum Iv_{/11}$) is obvious, along withthe need for updating the knowledge acquired on the basis of standardized units that appeared before and after 2008. According to the shown tendencies and frequencies of development, established original mathematical relations, presented trend lines, as well as clustering, the periods for updating the knowledge in the PDCA have been clearly defined. What follows is the publication of new research and significant aspect of the solutions to the practical "problem" (which are "theoretical" at first sight - standardized aspect for the needs of everyday practice).

ACKNOWLEDGMENT

The work presented here was supported by the Serbian Ministry of Education and Science – project III 44006, http://www.mi.sanu.ac.rs/projects/projects.htm#Interdisciplinary and project III 41007

REFERENCES

[1] ISS - Institute for Standardization of Serbia (2014), Advanced search: http://www.iss.rs/, http://www.iss.rs/standard/advance_search.php (accessed: 31.12.2014)

[2] ISO (2014), ISO Store, http://www.iso.org/iso/home/store.htm, health care technology, 31. 12. 2014.

[3] Crosby B. Philip (1980), Quality is free, Mc Grew -Hill, Inc. (translate: Qualitass International, Beograd, page 259)

[4] Mićić Ž., Mićić M., Blagojević M., (2013), ICT innovations at the platform of standardization for knowledge quality in PDCA, Computer Standards & Interfaces, 36(1), 231-243.

[5] Ž. Mićić and M. Demić, "Knowledge standardization in road vehicle engineering", TTEM, Vol. 7 (3), pp. 1281-1288 (2012)

[6] Ž. Mićić and N. Stanković, "Knowledge trends at the standardization platform of environment, health protection and safety", 16th International Conference ICDQM-2013, June 27-28. Belgrade, page 519-529 (2013)

[7] Ž. Mićić, M. Vujičić and V. Lazarević, "Analysis of Knowledge Base Units within Standardized Electrical Engineering Subfields", Acta Polytechnica Hungarica, Vol. 11, No. 2, 41-60 (2014)

[8] Ž. Mićić, M. Blagojević and M. Mićić, "Innovation and knowledge trends through standardisation of IT applications", Computer Standards & Interfaces, 36(2), 423-434 (2014)

[9] Ž. Mićić and N. Stanković, "Knowledge and innovations trends in metallurgy subfields within standardization platform", Metal. Int. XVIII (8) 154 – 160 (2013)

[10] Ž. Mićić, S. Petrović, Knowledge innovation trends on a standardization platform – in parallel: Civil Engineering and Railway Engineering, VIII International Conference "Heavy Machinery-HM 2014", Zlatibor, 25-28 June 2014, Proceedings: Session C: Civil engineering and materials page 25-33

[11] T. A. Walasek, Z.Kucharczyk, D. Morawska-Walasek, "Assuring quality of an e-learning project through the PDCA approach", Archives of Materials Scinence and Engineering, Volume 48, Issue 1, 56-61 (2011)

[12] N. Igari, "How to successfully promote ICT usage: A comparative analysis of Denmark and Japan", Telematics and Informatics, (2012) http://dx.doi.org.proxy.kobson.nb.rs:2048/10.1016/j.tele.2012.10.001

[13] K. Luu and G. J. Freeman, „An analysis of the relationship between information and communication technology (ICT) and scientific literacy in Canada and Australia", Computer & Education, Volume 56, Issue 4, 1072–1082 (2011)

Use of Geographic information systems in analysis of telecommunication market

Mirjana Kranjac*, Uroš Sikimić**, Đorđije Dupljanin***, Srđan Tomić****
* Faculty of technical sciences Novi Sad/Department for transport, Novi Sad, Serbia
** Politechnico di Milan/Department for management, Milan, Italy
*** Faculty of technical sciences Novi Sad/Department for transport, Novi Sad, Serbia
**** Fakultet za inženjerski menadžment, Beograd, Serbia
mirjana.kranjac@uns.ac.rs, uros_sikimic@hotmail.com, ddjordjije@gmail.com, srdjan.tomic@fim.rs

***Abstract*—The paper presents utilization of the Geographic information systems (GIS) as a tool for analyzing the telecommunication market. The location of points of sale of different telecommunication operators is in focus of research. Visualization which is attainable through the application of GIS gives results that can be used to create better distribution of points of sale for mobile operators.**

I. INTRODUCTION

The main goal of the paper is to find how the geographic information systems could be used in analysis of telecommunication market. The specific goal is to find how the GIS could be used for efficient location of points of sale for mobile operators. Task of the paper is to animalize overlapping of zones around points of sale of two mobile operators in Novi Sad by using GIS as software application for visualization. The result of analysis will give some recommendations for the relocation of points of sale which could give better economical effects and principles of future locations of new points of sale. Conclusion is that GIS can be powerful support for visualization and decision making about market configuration, simple and fast.

II. LITERATURE OVERVIEW

The geographic information systems are family of software for visualization of geo referenced data. They could be used for different analysis which offers visualized solutions. Geographic referenced data which are visualized are more acceptable for human use. Geographic location is the element that distinguishes geographic information from all other types of information. Without location, data are termed to be non-spatial and would have little value within a GIS. Location is, thus, the basis for many benefits of GIS: the ability to map, the ability to measure distances and the ability to tie different kinds of information together because they refer to the same place.

The paper [1] communicates the richness and diversity of GIS in an accessible format. It reinforces the view of GIS as a gateway to science and problem solving. Reference [4] gives a contextualized professional development approach for geographic information technologies as a project-based science.

Innovative approach for analysis by using GIS is described in articles of reference [5].

Geographic information systems (GIS) are presented and explained as decision support in the papers referred in [7, 8].

Mobil operators are drivers of big income for telecomm companies. Reference [2] is making analysis of their income drivers and is discussing many tools and market possibilities of GIS as a useful software to reach valuable decisions.

There are many telecommunication operators' challenges and roles. Authors in reference [3] are creating some sceneries for perspective of mobile commerce value chain by using information technologies but without visualization of results.

Some European economists are giving a good presentation of competition in two sided markets of two groups of operators as referred in [6].

III. GEOGRAPHIC INFORMATION SYSTEMS

Due to the increasing development of science and technology and development of human society, the importance of information becomes larger and extends from the information about idea to the information that can cause change of the space that is surrounding us. More and more, information is deeply connected to the location within the space. Therefore, there are developed different survey methods to collect spatial data, and the different ways of collecting, storing and sorting them, and better analyzing according to various criteria. The base is visualization which people are familiar with and which offers fast recognition of the situation or problem and easier solving them and visual presentation of the solution.

The spatial data describe the location, shape and orientation of objects in space. They are known as geospatial data. Geospatial data can describe the characteristics of many different types of objects on the surface. These objects can be tangible, physical things such as an office building, landscape or abstract forms such as the imaginary line that marks the political

boundary between the states.

From the moment when first started collecting spatial data and displaying them on maps, there is a tendency for them to be systematized and made available. Over a long historical period, the most effective way to display spatial information was analogous to the map. Map was the forerunner of the original spatial database, the original spatial information systems, and in some ways a forerunner of spatial data infrastructure. Map is important to display the spatial distribution, connectivity and interaction of objects and phenomena, as well as qualitative and quantitative alteration of the condition over time.

Technologies of GIS are the "main culprit" for the change. Geographic information integrated into other products and software applications have become mass-market product. Forerunner of spatial data infrastructure in today's terms it is probably the concept of integrated mapping of different thematic layers of data from the sixties. Thanks to information and communication technologies, conventional way of presenting information about the space is the past. Today, spatial data is usually collected, stored, processed, analyzed and presented in digital form through a number of applications. Spatial data types provide fundamental abstraction mechanisms for modeling the geometric structures of spatial data, their attributes, operations on them, and the relations between them.

"A GIS is a collection of software that allows you to create, query and analyze geospatial data. Geospatial data refers to information about the geographic location of an entity. This often involves the use of a geographic coordinate, like a latitude or longitude value. Spatial data is another commonly used term, as are: geographic data, GIS data, map data, location data, coordinate data and spatial geometry data." [9] .

GIS is a system for managing spatial data and their associated properties. In the strictest sense, it is a computer system capable of integrating, storing, editing, analyzing and displaying geographic information. In a broader sense, GIS is a "smart card" that gives users the ability to set interactive queries (user created surveys), analyze spatial information, and edit data. The technology of geographic information systems can be used for scientific investigations, resource management, property management, development planning, regional planning, mapping and infrastructure planning. GIS is often used for the purposes of market research, then in geology, civil engineering, but also in all areas that use data related to the map.

A. Work in GIS

Spatial data on artificial and natural objects are subject of GIS presentation and analyzing. These are, for example, infrastructure facilities, housing, sports and other buildings, woods areas, river flows, land use, elevation terrain data, geological data, traffic lines, political borders etc. They are called features in GIS. Each feature has its attributes. If feature is a mobile telecomm operator in Serbia, than the attributes could be: number of users per towns, number of different services per area, income per town, mobility of users per area ...

GIS data must be georeferenced and presented within spatial areas.

Spatial maps can be configured in the form of:

- Grid or raster and
- Vector.

Grid is consisting of rows and columns of cells, called pixels, where each cell has a single, particular, digital value. In the case of images, the numerical value represents the number of colors (colors are coded with numbers). The pixel value can not only present the color but also can represent the spatial data. For example, in some areas there have been excesses in the form of emission of toxic gases. Raster that shows the amount of pollution, consisting of pixels whose numerical values carry spatial information on the concentration of toxic substances in the air in each fraction of the space.

Displaying information in vector form refers to the geometry of the shape (length, height, and shape), either in terms of point, line or polygon entities and their spatial position (the position of the coordinate system).

Example: noise can be spatially displayed in the form of irregular polygons (polygons entity) and non-spatial data tied to this can be information about the types of trees, their numbers, percentages, different types, etc.

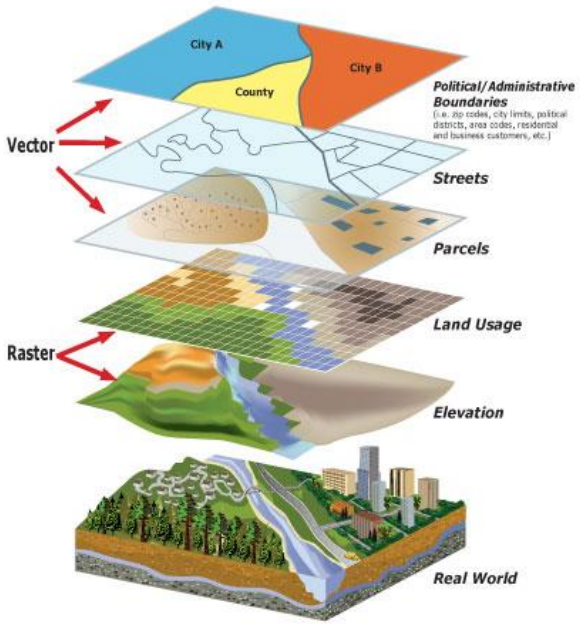


Figure 1: Example of vector and raster layers in GIS

B. Application of geographic information systems

GIS is an essential tool in all areas of design, planning, management and analysis. Around the world, used by electrical, mechanical and civil engineers, architects, bankers and economists, journalists, teachers, geodetic engineers, environmentalists, politicians, criminologists,

planners, health professionals and others.

Business people see the world as a collection of information about sales, customers, warehouses, demographic profiles and much more. The basis for all of this information is the address, sales region, or transport routes of delivery which can all be displayed and interactively operated on the map.

The planning and engineering tasks that can be easily solved with the help of GIS must go through the following steps:

- defining a problem
- collecting data (georeferenced)
- visualization,
- analysis,
- supervision and monitoring for a long-term period,
- corrections,
- new solutions.

Competitive pressure and new legislation causes an efficient and responsible management. This requires access to information based on a geographically distributed elements and operations. In today's world, competitive, successful management requires the maximum of all the resources, people, equipment and information. Using GIS to integrate geographical with other relevant data, gives the system fully equipped for this task.

C. Analysis with GIS

Not only presentation of existing geo referenced data is advantage of GIS. GIS also enables to put them into different layers and to overlap them. This makes possible to follow the trends of some features, for example increasing rate of citizens' number in a state within the timeline. There are many other GIS analyzing tools. Some of them are described:

Extract

• Clip: Extracts input features that overlay the clip features. It is as feature is like a cookie cutter, selecting only the part of the data set to be clipped that are within its boundaries.

• Select: Extracts data based on attributes. For example, if there is a map of all countries in the world, that contains a field giving each country’s continent. The select utility can be used to select only those countries with the continent field equal to “Europe.”

• Table Select: Extracts selected table records or features from an input table or table view and stores them in a new output table.

Overlay

Erase: Creates a feature class by overlaying the input features with polygons of the erase feature. Only those portions of the input features falling outside the polygons of Erase features are copied to the outside feature class.

Intersect: Computes a geometric intersection of input features. Features or potions of features which overlap in all layers will be written to the output feature class.

Spatial join: Creates a table join in which fields from one layers attribute table are appended to another layers attribute table based on the relative locations of the features in the two layers.

Union: Computes a geometric intersection of the Input features. All features will be written to the output features class with the attributes from the input features, which it overlaps.

Proximity

Buffer: Creates buffer polygons to a specified distance around the input feature which presents zone. Due to different demands GIS can make intersection, revised intersection or different.

IV. USE OF GEOGRAPHIC INFORMATION SYSTEM IN ANALYSIS OF TELECOMMUNICATION MARKET

This chapter describes investigation of locations of sales points of mobile operators VIP and Telenor in Novi Sad in order to increase profitability of them. This case is an example how GIS could be used in analysis of telecommunication market. The task was to find the best way to see interaction of sales areas surrounding the points of sale of two competing mobile operators: VIP and Telenor in Novi Sad.

To do this the following steps were performed:

1. Definition of the problem:

-To increase profitability of points of sale

2. To solve the defined problem, it was decided to find interaction between areas about location of point of sales of two operators and to analyze their interaction in two ways:

- to find locations which are covered by only one operator

- to find locations which are covered by the both operators

To do the previously tasks by using GIS it was necessary to perform the following:

1.To collect locations with postal addresses of points of sale for VIP and Telenor.

2.To present these locations within the map of Novi Sad in two layers: one for Telenor and second for VIP..

3. To create areas of customers gravitation. It was done by creation of new two layers which present areas surrounding locations where customers are oriented towards certain point of sale. These areas are called buffer zones in GIS:

-for Telenor and

-for VIP.

4. To create new layer which presents gravitation areas which are covered by only one operator (or first or second).

5. To create new layer which presents gravitation areas which are covered by the both operators (their intersections).

6. To discuss results and give conclusion.

The use of geo referenced data, their visual presentation and analysis based on GIS tools should have

given solutions for better profitability of both operators. The authors tried to create a tool for better localization of sale points having in mind the locations of competition. Figure 1 shows location of sale points of operator Telenor and Figure 2 of operator VIP.

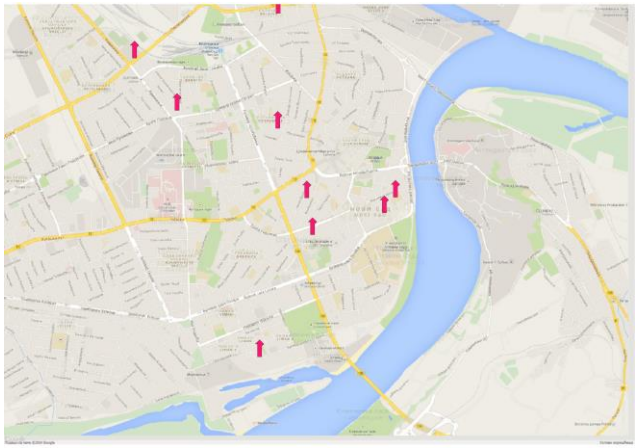


Figure 1. Locations of Telenor sale points

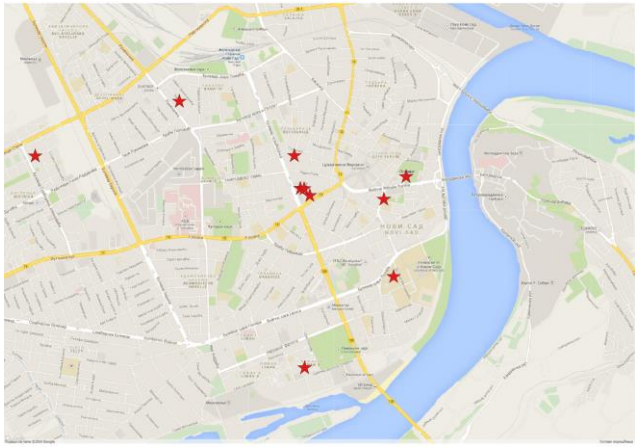


Figure 2. Locations of VIP sale points

By using buffer analyzing tool of GIS authors created Figure 3. which presents areas covered by sale offices of VIP.

Analysis are presented in Figure 4 and Figure 5. Figure 4 presents areas covered by only one operator, VIP or Telenor. Figure 5 presents locations which are covered by the both operators.

Figure 3. Locations of VIP sale points with areas surrounding them - zones buffers

Figure 4. Areas which are covered by only one operator

Figure 5. Areas which are covered by the both operators

Use of GIS as tool for analyzing mobile operators market gave visible results which could be directly implemented and which contributions could be measured by economical indicators. The paper presents small contribution in analysis of areas of sale points. Results show that sale areas of both operators are in many cases overlapping each other. Also, there are many areas covered by the both operators. Small areas are covered by only one operator and large space is not within sale areas

of any of them. This indicates lack of spatial research of areas of sale in telecomm market. Visualization which is attainable through the application of GIS gives significant results that can be used to create better distribution of points of sale for mobile operators. Recommendations for the relocation of points of sale which could give better economical effects are:

- stay located in the same zones as concurrency if profit is satisfactory, if it is not, points of sale should be moved
- try to move points of sale into "empty" areas. Select those "empty" areas where the density of citizens is the highest. This could be done by adding layers which presents density of citizens what would assist to select the right locations.

To locate new points of sale should be done the following:

- create layer with existing distribution points of operators
- create layers with density of citizens according to spatial zones
- create other layers with data of interest
- create decision rules
- overlap layers and find solutions.

V. CONCLUSION

Geographic information system should be implemented into the analysis of telecommunication market because of its visualization possibilities and possibility to overlap different layers. It will enable to do cross cutting analysis of different factors, to follow the behavior of other attributes at the market and to investigate effects of their own changes. Conclusion is that GIS can be powerful support for visualization and decision making about market configuration which will bring significant economical benefits for the actors of telecommunication market.

REFERENCES

[1] P. A. Longley, M. F. Goodchild, D. J. Maguire, and D. W. Rhind, *"Geographic information systems and science",* John Wiley, Chichester, 2001.

[2] A. Wilder, JD. Brinkerhoff, TM. Higgins, "Geographic information technologies+ project-based science: A contextualized professional development approach" - *Journal of Geography*, Taylor & Francis, 2003, pp. 123-132

[3] G. C. Moore, I. Benbasat, "Development of an instrument to measure the perceptions of adopting an information technology innovation". *Information Systems Research 2(3),* 1991, pp. 192-222.

[4] RS Bednarz, SW Bednarz *"The importance of spatial thinking in an uncertain world Geospatial Technologies and Homeland"*, 2008 – Springer.

[5] T. Randall, Cameron, Ch.,.B., J Churchill, Brian Baetz, "Geographic information system (GIS) based decision support for neighbourhood traffic calming" *Canadian Journal of Civil Engineering,* 2005, 32(1): 86-98, 10.1139/l04-085

[6] P. Roma, G. Perrone, F. Valenti, "An empirical analysis of revenue drivers in the mobile", *Journal of Management Information Systems* 18(2), , 2008, pp. 89-106.

[7] Y. Kuo, C. Yu, "3G Telecommunication operators' challenges and roles: A perspective of mobile commerce value chain" *Technovation 26* (12), 2006, pp. 1347-1356.

[8] Rochet J. C., J. Tirole, "Platform competition in two sided markets", *Journal of the European Economic Association* 1(4), 2003. Pp. 990–1029.

[9] Linux journal, ISSN 1075-3583, Belltown Media Houston, http://dl.acm.org/citation.cfm?id=J508, visited at 01.08.2014.

New Regulatory Approach in ICT Sector

Branka Mikavica*, Nataša Gospić*
* University of Belgrade, Faculty of Transport and Traffic Engineering, Serbia
b.mikavica@sf.bg.ac.rs, n.gospic@sf.bg.ac.rs

***Abstract*—The information and communication technology (ICT) sector continues to experience significant changes. Choosing and adopting appropriate regulatory tools to respond to new market behaviours is increasingly complex for regulators in converged environment. In such dynamic environment, new regulatory approach, a fourth generation of regulation, is essential to enhance digital communications. The paper presents new regulatory trends emerging in ICT sector with particular to Serbian ICT market situation. Challenges that ICT regulators need to deal with as well as new opportunities arising from the growing interconnected network environment are also shown in this paper.**

I. INTRODUCTION

The ICT sector remains one of the most rapidly evolving industry segments. The ever-expanding digital world influences almost all aspects of modern life. Nowadays, access to online services is essential in order to find a job, receive a salary, learn and make individual and business decisions. Overall goal is to bring ICT close enough to everyone. In achievement of this goal, innovation, investment and protection of the customer rights by encouraging the development of modern and effective regulatory tools are necessary. ICT regulators recognize that in such dynamic environment, new regulatory approach, so called a fourth generation of regulation, is required in order to enhance development of ICT sector.

The paper is organized as follows. After introductory remarks, the second section gives overview of global growth of ICT sector, as well as Serbian ICT market situation. Section III presents new regulatory trends and main and main issues for the fourth generation of regulation are described. In Section IV economic impact of ICT sector is presented. Section IV analyses potential challenges and opportunities of the new regulatory framework. Serbian ICT sector regulation is described. Concluding remarks are given in Section VI.

II. GROWTH OF THE ICT MARKET

Mobile broadband networks are being developed at an increasing pace. About 50 per cent of the world's population was covered by a 3G network in 2013 [1]. The migration to Long-Term Evolution (LTE) technology takes place much faster than did the earlier migration from 2G to 3G networks. According to the GSM Association (GSMA), commercial LTE networks were operating in 88 countries in 2013, up from 14 in just three years. The Global mobile Suppliers Association (GSA) puts that number at 101 countries. Ericsson estimates that by 2019, 65 per cent of the world's population will be covered by LTE, an increase from just 10 per cent in 2012 [1]. Mobile broadband (3G and 4G) shows the highest growth rate of any ICT, growing almost 20% during 2014. Additionally, LTE-Advanced is now commercially deployed on 9 networks in 7 countries worldwide [1].

The apps market involves new communications behaviours, new business models and includes a redefinition of the customer role. New apps and services, available on mobile connected devices, are offered to customers in order to inform them, play games, share files, exchange instant messages and videos, watch movies etc. The impact of all Internet-connected devices, apps and services on communications networks is enormous, making bright future for equipment vendors, manufacturers and apps providers.

Cloud services and data analytics (also known as "big data") set additional strain on networks. Operators and service providers work intensively in order to identify strategies to cope with the ever-increasing traffic expansion.

Developing nationwide broadband infrastructure remains a key goal in most countries' digital agendas and plans. Mapping the deployment of fibre transmission capacity require public funding due to a lack of private-sector economic viability. Although great efforts have been made to increase international connectivity, many countries face challenges in deploying and expanding next generation networks in order to support the ongoing growth in data traffic [1].

Today, xDigital Subscriber Line (xDSL) still accounts for over half or more than five out of every ten fixed broadband lines, with fibre optic Fiber to the X (FTTx) accounting for around a quarter of the total market for fixed broadband. Fibre is growing slowly, but permanently – Fiber to the Home/Fiber to the Building (FTTH/FTTB) account for over a fifth of all connected households in just nine countries worldwide. High household penetration of broadband is a key indicator of market maturity.

The Internet of Things is another strong demand generator for broadband. Embedding technology into everyday environment is likely to cause major social changes, as it will become more possible to track people's movements, activity, interactions and interests, all of which raise major issues regarding privacy, security and personal protection. Since broadband environment is expanding continuously and involves non-traditional ICT and Internet players as well as providers from other sectors, reassessment of ICT regulation in order to bring more flexible approach to regulating issues at different levels is necessary [1].

One way to monitor progress in ICT developments in developed and developing countries, as well, and to measure the evolution of the global digital divide is ICT Development Index (IDI), introduced by International Telecommunication Union (ITU). It presents a composite

index that combines 11 indicators into one benchmark value on a scale from 0 to 10. IDI consist of three sub-indexes: the access sub-index, the use sub-index and the skills sub-index. The difference between the values of IDI between developed and developing countries is large. IDI values in developed countries are on average twice as high in comparison with developing countries. The developing countries are very heterogeneous in the terms of IDI. There is a great difference between IDI values for the highest and the lowest country. Developed countries are more homogenous regarding these values. Also, it can be noticed, that the highest growth is recorded in developing countries, not only on the IDI overall, but on both the access and use sub-indices, as well.

Last measures show that all countries in European region, with exception of Albania, have IDI values higher than 4.77, which is global average score for IDI. Values of IDI for Serbia show prominent increase during last five years [2, 3, 4].

Another way to define and evaluate market growth is Network Readiness Index (NRI). The World Economic Forum defines NRI as a nation's or community's degree of preparation to participate in and benefit from ICT development. This framework evaluates nation's challenges and opportunities regarding ICT. Important stakeholders regarding development and use of ICT are individuals, businesses and governments who realize their roles in a general macroeconomic and regulatory environment. The degree of usage of ICT by stakeholders is linked to their degrees of readiness and capability benefit from ICT. Component indexes are: environment, readiness of nation and usage.

The environment component index measures potential of the environment that a country provides for the development of ICT. Sub-indexes that encompass environment are market, political/regulatory and infrastructure. The readiness of a nation presents a measure of the capability to support the potential of ICT. Sub-indexes that present a measure of readiness are business, individual and government readiness. The usage component measures the degree of usage of ICT by the principal stakeholders mentioned above. This component presents an indication of changes in behaviours, lifestyles, and other economic and non-economic benefits arising as a result of adoption of ICT. Sub-indexes used for measuring usage are individual usage, business usage and government usage. Measurement in the last five years shows that Sweden, Finland and Singapore have the highest NRI (Sweden: 5.65, 5.60, 5.94, 5.91, 5.93; Finland: 5.44, 5.43, 5.81, 6.04; Singapore: 5.64, 5.59, 5.86, 5.96, 5.97) [5, 6, 7, 8].

The ICT market in Serbia has been experiencing expansion for years, particularly in the number and structure of Internet connections and the total revenues from the Internet service provision. Positive growth trend maintained in 2013, with the total number of broadband users (without accounting for 3G network users) equals 99% of all Internet connections, which is approximately 8% more than in 2012 [9]. ADSL access represented the dominant Internet connection in 2013, accounting for 47% of all broadband connections (without 3G network subscribers) [9]. In addition to the ADSL, other means available for the Internet access were cable modem, which is another service provided by the CATV operators, directly, via Ethernet, via optical cable, by means of wireless access in the 2.4 GHz and 5.8 GHz unlicensed frequency bands, less often using the 3.4-3.6 GHz frequency band, as well as via mobile operators' network (either via cell phone, or by means of special modems) [9].

The growth of the ICT market is influenced by the increased number of users as well as by the total revenues from the Internet service provisioning during past years. However, a certain slowdown in growth is notable in 2013. in comparison with the previous period as a result of market saturation and general economic trends [9].

Fig. 1 shows Serbian NRI in the last five years [5, 6, 7, 8]. It's notable that network readiness increases, but ranking varies from. Thus, in 2010, Serbian NRI has score 3.51 taking rank 84; in 2011. NRI score has value of 3.52 at rank of 93; in 2012. NRI score is 3.64 and rank 85; in 2013. NRI score is 3.70 and rank 87; and the best recorded score is measured in 2014, where NRI score equals 3.88 and rank is 80 [5, 6, 7, 8].

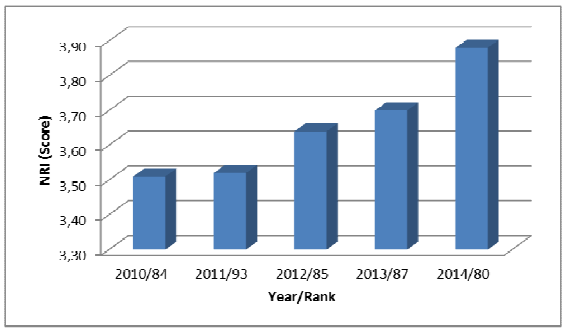


Figure 1. Serbian NRI in the last five years

Considering neighbouring countries, latest measures in [8] show that only Albania has lower NRI (3.66) in 2014, in comparison with Serbia (3.88). The highest NRI in Serbian neighbourhood has Croatia (4.34), as depicted in Fig. 2.

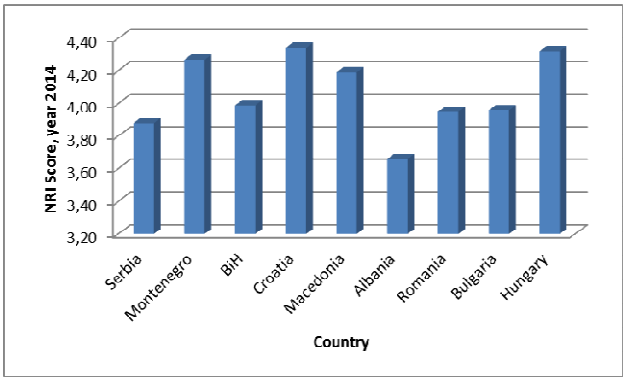


Figure 2. NRI for Serbian neighbouring countries in 2014.

III. Challenges and Opportunities in ICT – Regulatory Aspect

As technology and media continues to evolve, regulatory challenges that ICT sector need to cope with increase rapidly. This evolving environment offers many opportunities for operators. However, the ability of understanding and managing the corresponding risks are essential in taking those advantages. In such environment, network operators need to ensure that their risk management of the business keeps pace. Hence, major

challenges for operators from regulatory perspective are failure to realize new roles in telecommunication market, lack of regulatory certainty on new market structure, new imperatives in privacy, security, and data integrity [10].

Reduced macroeconomic uncertainty is positive climate for operators. However, structural regulatory pressures on core service areas to increased competition from over-the-top (OTT) players mean that market conditions remain challenging. Operating environments vary significantly between regions, causing divergence of telecom industry performances. The main consequence of this divergence in regional performance is lack of confidence in the sector's fundamental attributes.

Implementation of LTE networks and increasing customer demand causes that network capital expenditure remains at elevated level. This trend supports current regulatory models where competition presents catalyst for higher quality services. A combination of price deflation driven by competition from OTT providers and adjacent market players mean that telecommunication market conditions for operators remain highly challenging [10].

In emerging markets, shortages in spectrum and low prices require consideration of more rational market structures, either through consolidation or new wholesale mobile broadband networks. Regulators already recognize the need to reform existing rule to enforce development of ICT sector. However, market consolidation remains unclear. Pro-competition and pro-investment policies need to be balanced by regulator. At the same time, operators need to seize the initiative by prioritizing shared market positions and re-examination relative merits of in-market consolidation, in order to encourage progress. Market-structure related issues such as wholesale mobile network provision require attention both regulators and operators. Other relevant issues include provisions around spectrum release, licensing and sharing. At the same time, operator's willingness to innovate their business models depends on new relationships with other entities [10].

ICT is one of the most important sources of new opportunities to encourage innovation and to intensify economic and social improvement, for both advanced and emerging economies. The risk of developing this kind of data-driven policy and regulation comes from exposure of private data to either companies or governments.

The uncertainty and concern around data privacy and security become severe issue expanding into new areas such as data sovereignty and internet governance. This issue has contributed to highly ambivalent attitudes among customers over access to reuse of their personal data. Some segments of the customers are willing to release personal identification and usage data on an anonymized basis. Still, many customers are convinced that information on customer purchasing and related behaviour benefit companies gathering big data more than end users themselves. Operators can reverse declining levels of trust among end users by enhancing privacy and security features to their service propositions. At the same time, proactive attitude is required on privacy and security issues with partners and policy-makers so that new demands for data sovereignty, personal data privacy and cybersecurity can be complied with in the long term.

Some new approaches to regulation and technology that contribute personal privacy protection from misuse are being developed. The focus is on the protection of individuals with regard to the processing of personal data and on the free movement of such data [11].

Operators are considered as having a natural competitive advantage in the big data field in comparison with other industry areas due to their legacy of strong customer, network and product information assets. The scope and diversity of this information presents additional challenges for telecommunication operators, since they use big data in the process of creating values within and beyond their organizations. Strategies concerning big data are important issue at many leading operators. There are numerous of challenges emerging in repurposing of various data, showing lack of cooperation within organization, lack of leadership understanding and commitment around importance of data, and fragmentation of data sources all hindering progress. All these challenges have led to the fact that proportion of budgets projected to be devoted to big-data solutions is expending significantly in the coming period. However, there is no guarantee that these actions will generate values from big data. Long term benefits require carefully represented strategies to be balanced. Having this goal set, value opportunities from big data must be defined and prioritized.

IV. ICT Sector Economic Impact

The ICT sector is major contributor to economic development. Generation of revenues and securing investment in telecommunications is increasingly gaining in importance. Revenues from telecommunication sector in developed countries are under pressure from several reasons. These markets are highly competitive. In the mobile service segment many are close to saturation. However, number of customers is still growing. In addition to tighter customer budgets as a consequence of economic crisis, operators need to deal with revenue pressure from applications, which diminish traditional revenue flows [5].

Mobile sector is main source of revenues in developing countries, approximately 62% of total telecommunication revenues, and this portion continues to grow. Renewed investment is essential to satisfy requirements of advanced services, especially broadband. Progress would never be possible without major investment in telecommunication networks. Nowadays, investment is needed in improvement of existing services and their upgrade to broadband, but also to enable the access to more customers. Hence, monitoring of investment is essential issue for policy-makers [5].

ICT sector is infrastructure-intensive business. It requires large-scale and long-term capital outlays. For realisation and spreading of incomes earned on investment several years are needed. In a highly competitive environment, such as ICT, renewed investment is of great importance in order to meet the requirements of advanced services, including applications demanding large bandwidth and convergent services, for fixed broadband, and mobile broadband services, as well. One of the ways to measure investments in fixed assets needed to support development of ICT, regardless of the origin of capital, whether is domestic or foreign, public or private, is monitoring of data on capital expenditure. Results show that investment has declined in developed countries and increasing trend is notable in developing countries. The

ratio between capital expenditure is below 20% in most developed countries, and above 20% in the majority of developing countries. This emphasize the fact that most developed countries in terms of ICT deployment require relatively low levels of investment relative to revenue generated by telecommunication services. On the other hand, developing countries require more significant relative investment in order to enforce growth [5].

Data on foreign direct investment in ICT sector raise question on the cross-border movement of finance capital and on the extent of business internationalization in this sector. Recorded results show that after economic crisis, foreign direct investment in ICT declined significantly, but developing countries were less affected. Foreign investors agreed deals in developing countries, which present better economic prospects and are recognized as important sources of revenue growth. Developed countries remain the leading source of financing for foreign direct investment, but developing countries are getting an increasing role [5].

Since economic crisis affected ICT sector, operators feel pressure to provide services at optimal low-cost performance. Ensuring customer privacy is necessary step in achieving business success. On the other hand, only the market players that can reach all available customer data are in position to actually accomplish that goal. In this environment risk managers need to leverage traditional loss prevention tools in more sophisticated ways. It consists of new methods to transform their functions into profit centres in order to enable positive contribution to the company development. Due to the analysis of many consulting agencies, three leading success factors are: increasing of revenues and margins that can be achieved by share-of-wallet and greater market share, identification of new revenue sources such as data monetization, and improvement of operational efficiency [12].

V. NEW POLICY AND REGULATORY TRENDS

The evolution of telecommunication regulation can be described through several phases. The first generation presents monopoly utilities, public or privately owned, that were closely managed. The intent was to encourage improvements in efficiency and service. In this situation, regulation had task to simulate the desired effects of competition. The second generation is characterized by partial privatization and licensing of competing infrastructure providers. This regulatory phase is focused on balancing of opening up the access to incumbent's network with the need of protection government infrastructure investments and ongoing shareholdings. The third generation brought full privatization. Regulation is shifted toward protecting competition in service and content delivery, with an increasing need for customer protection. Due to market and technology development, government policy-makers face even greater need to ensure access to digital infrastructures, primarily to fixed and mobile access networks. Broadband networks are more becoming non-optional utilities, even rights, whose availability and performance impact every aspect of the economy and societal development [14].

Choosing and adopting appropriate regulatory tools to respond to new market demands and the growing need for customer protection is increasingly complex for regulators in converged networks environment. Considering these various issues, regulators must be aware of international context within which they operate. In the mobile sector field, international allocations are realized in the development of regional band plans that guide spectrum usage. On the fixed network field regulators are striving in improvement of Internet access with cost reduction. The goal is to ensure fair and effective traffic management that balances customer demand and needs of network operators and content/service providers. In order to insure customer rights, regulators need to address multiple issues such as securing privacy and data protection in a cloud environment, and raising user awareness on the appropriate use and impact of shared content [13].

Analysing new trends in ICT market and all challenges and opportunities from its development, it can be noticed that stakes for regulators have never been higher. Regulators in this new generation of regulation must be able to oversee an increased range of services, delivered over multiple broadband and converged networks that create the digital ecosystem. Regulators also must protect customers from inappropriate content, faulty billing and fraudulent online activities. Another important issue regarding the fourth generation of regulation is involvement not only in economic necessity of creating affordable access, but also in the attendant social opportunities and challenges arising from better-connected communities. This is primarily related to developing countries [14].

The evolution of regulator's role to the fourth generation regulation can be shown as a response to several critical issues coming from the changing of the environment. These issues, presented on the Fig. 3, stem largely from economic and social development realities and objectives set by government policy-makers. It is important to emphasize that these issues are additions to the more traditional tasks of regulators, which will become less important with the maturing of a competitive market place.

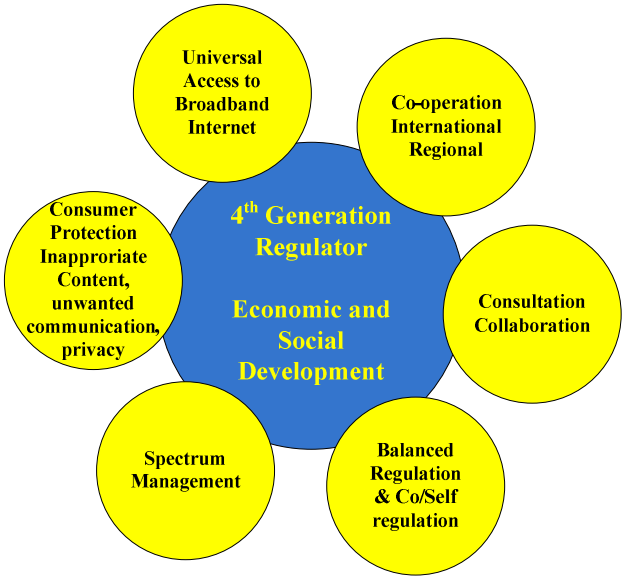


Figure 3. The evolving role of the regulator [14]

The main challenge for governments and regulators is how to enforce the private sector to cover as large a percentage of the population as possible, leaving only a small number of people to be connected by financial

subsidies. In this situation, the fourth generation regulator has a major role, working on a wide range of target groups in order to ensure universal broadband access. In developed countries and urban areas, this should be realized on a universal service basis, thus enabling every person or household to have access to broadband service. In developing countries and rural areas, the policy goal is to enable universal access, ensuring that each individual has access to broadband service somewhere in the community [14].

The source of success is in adopting policies that include broadband infrastructure. Effective program management involving all stakeholders is also needed.

International Telecommunication Union (ITU) highlighted the importance of allocation spectrum in an effective manner in order to meet increased demands for broadband wireless access [13].

Regulatory approaches to spectrum allocation and assignment for broadband communications has a direct impact on competition, costs and development speed. Spectrum should be allocated in a manner that maximizes its use and supports best economic and social outcomes. Where there is competition for spectrum access, auctions should be designed to achieve this optimal spectrum usage. Operators should be encouraged to improve efficiency and to maximize quality [14].

One option for regulators in the fourth generation of regulation may be to treat spectrum as a wholesale commodity, in fact, to charge a rent for its use but not requiring operators to lay out considerable capital expenditures to get licences. This approach can shift spectrum from a capital expenditure to an operational expenditure. As a result, operators would be able to deploy new infrastructure and to provide lower-cost services to customers. If operators were not effectively using the spectrum, they would lose the right of using it.

In the converged digital ecosystem, everything is interconnected. Standards and operational procedures are necessary to enable functionality of this system. Regulators need a common framework and forums where communities of interest can work together at a local, national, regional, and international level. A major role in facilitating such forums has the fourth generation regulator.

Customer protection activities are becoming more important with service convergence and the increased use of Internet. Many countries are adopting customer protection policies designed particularly for ICT customers, enforced either by regulator and/or a designated customer protection agency. Regulators need to collaborate with any and all agencies of interest in order to coordinate activities in the interest of customer protection. Also, regulators need to ensure that operators should clarify to their customers that complaints can be brought to independent regulatory entities.

Considering diversity nature of Internet, its openness and accessibility, some of undesirable elements of society have found ways to use the medium to commit fraud and other types of crimes. Government's response to these crimes is set of general laws, which aim is to protect citizens, especially children. However, government involvement in issues concerning content control is very sensitive, triggering questions about freedom of speech. Significant enhancement in content provision represents an important challenge to content regulation. The relevance of this issue is becoming even more important since large proportion of the content may originate in other jurisdictions. One of the appropriate ways for the regulators is to encourage operators to develop self- and co-regulation methods to cope with complaints from customers. These methods can even involve blocking and removing offensive material that caused customer complaint.

In many countries, as well in Serbia, regulators need to consult with stakeholders before publishing regulatory decisions, determinations or guidelines. Service convergence and development of the Internet brought a large number of these stakeholders who are much more broadly representative of society. The main role in advising and communicating with policy-makers in government ministries and offices belongs to the regulator.

In order to be effective, fourth-generation regulators need to satisfy characteristics such as:

- openness to ideas and approaches;
- flexibility to keep up with rapid changes in the market;
- business sense to work with operators;
- knowledge of financial aspects of the business;
- political agility and understanding to work with political leaders;
- the ability to offer policy guidance;
- the ability to develop appropriate regulations to implement public policy; and
- an understanding of consumer issues.

Also, regulator should be innovative and capable to achieve the vision and goals set by policy-makers, but still to remain within borders of law [14].

According to the Electronic Communications Act of Republic of Serbia, regulation of electronic communications are divided between the Government, the Ministry in charge and the regulatory body, the Regulatory Agency for Electronic Communications and Postal Services, RATEL. The Government determines policy in the area of electronic communications and adopts strategies and action plans for their implementation. The Ministry supervises the implementation of the law and associated bye-laws. Also, it adopts radio-frequency allocation plans and bye-laws related to technical conditions applicable to networks and equipment. RATEL is functionally and financially independent from all other state authorities. The supervision of operators is divided between RATEL and the Ministry. RATEL is authorised to verify operators' compliance with their obligations determined by the law, the implementation of bye-laws and decisions of the RATEL. To this end, it is authorised to request the necessary information from operators, to measure and test networks, services and equipment [15].

Activities in the process of Serbian ICT market liberalization can be divided into two phases. The first phase begins with the establishment of regulatory agency, RATEL, and lasts until 2010, when new law of electronic communication has been issued and marked a new phase in regulation. In the initial phase, despite great

obstruction, the first notable results have been achieved through licencing of three operators in mobile telephony. Further development is accomplished by opening the Internet market. Enforcing broadband access nowadays exist more than 230 providers in Serbian market. Similar approaches are deployed simultaneously on the content delivery market. Thus, there are more than 80 content operators. It can be noticed that major contribution of the first phase Serbian market regulation is diminishing of monopoly in all sectors, except in fixed telephony. There were some efforts to change this, but without any success. RATEL tried to eliminate the understatement of law by introduction of wireless fixed telephony (CDMA technology). Licences for this technology were granted to one existing and one new operator. However, this effort was too weak in comparison with monopoly in fixed telephony. Besides issues related to fixed telephony regulation, some additional issues emerged, such as prevention of potential monopolies, especially in the segment of relevant markets recognition and the application of multicriteria analysis for determination of operator with significant market power, whose prices are under a special regulation. Hence, prerequisite for final market liberalisation is achieved after issuing law of electronic communication in 2010. The major contribution of this law is abolition of monopoly in fixed telephony market. In addition to the ensuring regulatory conditions for deployment of new technologies, broadband communications and e-services, RATEL has announced five wholesale and four retail markets. Recognising a number of operators with significant market power, conditions for ex ante price control and cost based models are acquired. Thus, establishment of new monopoly is prevented [16]. Considering the need to follow the ICT sector development, RATEL has to be prepared for the transition toward the fourth generation of regulation.

VI. CONCLUSION

Regulators are major facilitators and partners in promoting development and social inclusion. One of the roles of regulator is sponsorship of public-private partnership among aid donors, governments, ministries and non-governmental organizations, especially in achieving universal access goals for rural, remote and un-served areas. In a converged ICT sector, the competition of operators and content providers is complex, especially if they report to different authorities on different issues. There is a need for softer, flexible regulation, free from bias. ITU is a leading promoter and argues for the transition to the fourth generation of regulation. In order to be effective, inter alia, the fourth-generation regulators need to be open for new ideas and new approaches, to show enough flexibility to keep up with significant changes in the market, to develop appropriate regulations to implement public policy. The regulators of this new generation of regulation differ from previous generation of regulator in the significance they place on the tendency of government social and economic policy goals, as well on the improvement of customer protection and broadband access.

ACKNOWLEDGMENT

This work is partially supported by Ministry of Education, Science and Technological Development of the Republic of Serbia under No. 36022.

REFERENCES

[1] A Report by the Broadband Commission, “The State of Broadband 2014: Broadband for all”, September 2014.

[2] ITU-D Report, “Measuring the Informational Society”, 2012.

[3] ITU-D Report, “Measuring the Informational Society”, 2013.

[4] ITU-D Report, “Measuring the Informational Society”, 2014.

[5] The Global Information Technology Report, “Transformations 2.0”, 2010-2011.

[6] The Global Information Technology Report, “Living in a Hyperconnected World”, 2012.

[7] The Global Information Technology Report, “Growth and Jobs in a Hyperconnected World”, 2013.

[8] The Global Information Technology Report, “Rewards and Risks of Big Data”, 2014.

[9] Republic of Serbia, Republic Agency for Electronic Communications, “An Overview of Telecom Market in the Republic of Serbia in 2013”, 2013.

[10] Ernst & Young Global Limited Report, “Top 10 Risks in Telecommunications 2014”, 2014.

[11] European Commission, “Proposal for a Regulation of the European Parliament and of the Council on the Protecion on Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (GDPR)”, Brussels, 2012.

[12] Experian Survey, “Trends in Telecommunication Industry 2014”, July 2014.

[13] ITU-D Summary, Trends in Telecommunication Reform 2013, “Transnational Aspects on Regulation in a Networked Society”, May 2013.

[14] ITU-D Special Edition, Trends in Telecommunication Reform, “4th Generation Regulation: Driving Digital Communications Ahead”, 2014.

[15] BDK Attorneys at Law, “Telecommunications in Serbia”, 2014.

[16] D. Malinic, V. Milicevic, “Finansijska stabilnost sektora telekomunikacija u Srbiji”, *Telekomunikacije*, vol. 10, November 2012, pp. 2-15.

CONTEXTUAL MODELING OF ICT PROJECTS FOR E-GOVERNMENT: THE CASE STUDY OF REPUBLIC OF SRPSKA

Milan Latinović*, Zora Konjović**
* Agency for Information Society of Republic of Srpska, Banja Luka, Republic of Srpska, Bosnia and Herzegovina
** University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Republic of Serbia
milan.latinovic@live.com, konjovic.zora@gmail.com

ABSTRACT - Public administration institutions continuously implement information technology in their business processes through individual ICT projects. To ensure the coordination of such projects, they should be described as to enable their comparisson and determination of their effects within the context of the broader goals of electronic public government. This paper proposes the metadata model and the conceptual model of evaluation aimed at evaluation of the projects' effects within the broader context of the development of electronic government by taking into account the project itself, participants in its realization, and the direct beneficiaries of its results.

Keywords: *ICT project, electronic government, methodology, semantics, context*

I. INTRODUCTION

Semantics (Greek sēmantikós) is the branch of linguistics and logic concerned with a meaning, while a *concept* is an idea or thought that corresponds to some distinct entity or class of entities, or to its essential features, or determines the application of a term, and thus plays a part in the use of reason or language [1]. Accordingly, the semantic concept of an entity represents a meaning of this entity, both for itself and for its environment.

Modelling semantic concept means definition of a clear purport of entities to which this concept applies and definition the interaction between these entities.

When considering the concept of the project, a number of papers define it in different ways. According to the PMBOK Guide [2], a project is defined as follows:

The project represents a temporary activity which is undertaken in order to create a unique product, service or result.

"Temporary activity" means that each project has a clearly defined beginning and the end.

"The unique product, service or result" may refer to:

- concrete product that has been created, which can be quantified and may represent the final product of the project or a partial component;
- the ability to perform a service, for example, business process that supports the production or distribution;
- delivery such as a document or outcome, (for example new knowledge that can define new processes or even formal documents such as rules or standards).

The definition of electronic government (e-government) used in the paper uses is the one given in [3]:

E-government is defined as *the use of information and communication technologies in public administrations combined with organisational change and new skills in order to improve public services and democratic processes and strengthen support to public policies.*

Modelling the context of ICT projects in e-government environment involves creation of metadata describing ICT projects in such a way to enable their monitoring, individually and collectively, with the aim of systematic improvement of e-government.

II. ENTITIES OF E-GOVERNMENT

Semantics modelling requires basic concepts to be defined together with relations among them.

From the perspective of this paper, basic concepts of e-government are the following entities: *Institution, ICT project, Service and Citizen.*

A. Institution

The main carrier of an ICT project is an institution, which is described by following metadata:

- institution name;
- unique identifier;
- institution authorities.

Within the context of achieving the broader goals of e-government, it is necessary to pay attention to the following:

- The institution can be independent, it might have a parent institution, and it may be superior to another institution.
- The institution can (temporarily or permanently) transfer jurisdiction to another institution in order to successfully implement project or some activity under that project.
- In case when more than one institution is working on same project(s) or implementing same process it is necessary to clearly define owner structure for

all aspects of project (processes, results, entry data and databases etc.).

B. ICT project

Based on the document [4], in the proposed ICT project model the following aspects of electronic administration development are defined as necessary aspects of projects monitoring:

- strategic aspects;
- project management aspects;
- technical and technological aspects.

Strategic aspects refer to the references to strategic documents which can influence project directly or indirectly. Also, strategic aspects may refer to the documents which will appear as a result of observed project, and will be used as direct or indirect reference for observed or any other project.

Project-management aspects are based on identifying stakeholders, identifying the applied methodology, risk assessment, defining the legal basis for the implementation of the project and the definition of business processes within the project.

Technical and technological aspects refer to the content of the technical documentation which should appear as a result of the project such as the hardware and software aspects, descriptions of databases, defining communication channels, aspects of interoperability, standards, and the like.

Register of ICT projects in public administration of Republic of Srpska (RegIKT) is the information system which supports metrics for defining level of development, identification of needs and coordination of activities related to e-government development. Information system RegIKT has been developed by using methodology described in [5]. According to this methodology, ICT project is defined by the following metadata:

- project name;
- level of government institution which is implementing this project;
- indication of whether the project is international;
- name of the institutions involved in the implementation of the public government (PG);
- contact information (internet site of the project, contact person);
- external partners;
- segments and complexity of the project;
- status of the project;
- start date of implementation;
- the duration of the project;
- value of the project and funding sources;
- contractor / executor / supplier;
- comment on the project; and
- additional information

Also, the methodology defines the conditions, such as:

- Multiple public administration authorities may be involved in the implementation of the project, but only one body can be responsible for the project (the main responsible institution for the project).
- The project value can be defined by the specific financial amount or described trough usage value.
- One project can be displayed as a number of smaller projects, separated by segments, as long as the total financial value of these projects is equal to the financial value of the parent project.

C. Service

Service actually represents a mechanism which enables certain institution to provide end users with its service. **Electronic service** is a service that is implemented by using ICT and, in most cases, it's realized through reengineering of one or more existing services.

Process represents a clearly defined algorithm/workflow that solves a particular problem.

Using the approach described in methodology used for ICT projects we are able to describe metadata for service as an entity.

Unique identifier of the service represents a primary key which is used to index services.

A list of previous versions of the service represents a set of unique identifiers that point to previous versions of service.

Service name is a descriptive name for the service.

Development environment of the service is the connection with the project / projects from whom / which this service is created or updated.

Related services are those services that are required for the operation of this service.

Owner of the service defines the institution that owns the service.

The service user identifies users who have the right to use the service. Each specific user is defined by the set of permissions for usage of the service (for example, right of access to databases, updates or reading data, etc.).

Interactivity of service describes the nature of the interaction that user achieves via service: access to content; electronic forms download; two-way communication without authentication; two-way communication with authentication; complete transaction service (authentication with transactions that have a financial impact).

Technical specification of service provides access to all documentation that represents technical specification for the service (i.e. database schemas, use cases, activity diagrams, network schemes, etc.).

Legislation of service provides access to all legislative documents, which affects the service and to which this service must comply.

Additional information provides access to information which is not covered in this metadata but is relevant for the given service.

D. Citizen

In the context of this paper, the entity Citizen does not represent a physical entity, but an abstraction that should provide information about the benefits that a citizen, as an individual, is achieving by implementation of specific ICT project. This information is provided by connecting *Citizen* entity with *ICT project* entity, where *Citizen* entity is

described with specific metadata, with semantics given in further text.

Citizen's profile represents the target group which includes all individual citizens who should enjoy the benefits from ICT project.

The intended benefit is a benefit that has been planned for the target group *Citizen's profile* trough implementation of ICT project.

The accomplished benefit is a benefit that has been achieved for the target group *Citizen's profile* trough implementation of ICT project.

If the entities *ICT project* and *Citizen* connect via metadata from *Services of the public administration* entity and *ICT project* entity, it is possible to provide such information for each service that is the subject of any ICT project.

III. CONTEXT OF ICT PROJECT

In order to define semantic model of ICT project, we contemplate the project through a series of identification and metadata shown in Figure 3.1.

UNIQUE IDENTIFIER OF ICT PROJECT		
PROJECT NAME		
LEVEL OF IMPLEMENTATION	INTERNATIONAL PROJECT	
PROJECT LEADER		
LIST OF PARTICIPANTS IN THE PROJECT	LIST OF EXTERNAL PARTNERS	
PUBLIC ADMINISTRATION SERVICES		
DEGREE OF PUBLIC ADMINISTRATION SERVICES		
PROJECT PHASES / LOTS LOT 1 LOT 2 - LOT N		
STATUS OF THE PROJECT		
START DATE	PLANNED COMPLETION DATE	**ACTUAL COMPLETION DATE**
SOURCES OF FUNDING		
TOTAL AMOUNT OF FINANCING PLANNED	**TOTAL AMOUNT OF FINANCING EXECUTED**	
SUBMITTED DOCUMENTATION		
NOTES		

Figure 3.1. – Context of ICT project

The semantics of metadata is described in sequel.

Unique identifier of the ICT project represents a unique value that identifies the given project. In database terminology, this is a primary key of this model. The practical implementation of this identifier can vary, depending on the environment that implements this model.

Project name represents the formal name of the ICT project.

Level of implementation represents level of public administration on which this project will apply. It defines if project will apply to the level of single town, region or whole country.

International project defines whether it is a project of international significance (i.e., project which includes partners from other countries). In the case of the acceding EU, this field may represent (or can be replaced by) a tag which indicates if the project is related to EU integration.

Project leader is a unique *Institution*, which is fully responsible for initialization, monitoring, project implementation and reporting.

List of participants in the project is a list of *Institution*(s) which participate in this project together with project leader.

List of external partners is a list of institutions participating in the project, but by their nature are not parts of the observed eGovernment system (i.e. Non-Government Organizations, private sector partners, external auditors and validators and the like).

Public Administration Services (one of the most important indicators of ICT project) represents the field of legal framework and practice for acceding EU. It can be defined according to the fields of the legal framework and practice of the EU [6]:

1. Free movement of goods;
2. Freedom of movement for workers;
3. Right of establishment and freedom to provide services;
4. Free movement of capital;
5. Public procurement;
6. Company law;
7. Intellectual property law;
8. Competition policy;
9. Financial services;
10. Information society and media;
11. Agriculture and rural development;
12. Food safety, veterinary and phytosanitary policy;
13. Fisheries;
14. Transport policy;
15. Energy;
16. Taxation;
17. Economic and monetary policy;
18. Statistics;
19. Social policy and employment;
20. Enterprise and industrial policy;
21. Trans-European networks;
22. Regional policy and coordination of structural instruments;
23. Judiciary and fundamental rights;
24. Justice, freedom and security;
25. Science and research;
26. Education and culture;
27. Environment;
28. Consumer and health protection;

29. Customs union;
30. External relations;
31. Foreign, security and defence policy;
32. Financial control;
33. Financial and budgetary provisions;
34. Institutions; and
35. Other issues

Degree of public administration services defines the degree of electronic services that will be achieved by implementing this project. This field represents a direct indicator of the progress of e-government, after the implementation of this project. This field is classified according to the classification defined in the paper "Strategic framework for the development of eGovernment and eServices in the world" [7], as follows:

1. presentation content;
2. access to web forms;
3. fulfilling the general forms;
4. transactions; and
5. Connection "full integration".

Presentation content – *Citizen* receives basic information about *Institution*, where *Institution* has no information about *Citizen*. This concept includes G2C and G2B concepts.

Access to web forms - *Citizen* receives access to *Institution* official forms, whereby *Institution* still has no information about *Citizen*. This concept includes G2C and G2B concepts.

Fulfilling the general patterns – non-authenticated two-way communication (forum, FAQ, polls, etc.). This concept includes G2C, G2B, B2G and C2G.

Transaction - *Institution* is aware of the identity of the *Citizen*, authentication is enabled as well as electronic transactions. This concept includes G2C, G2B, G2G, C2G and B2G.

Connection – *Institution*(s) are integrated into fully connected entities which are able to respond to the needs of *Citizen* by using interconnectedness of specific *Institution*(s), interoperable information infrastructure and connectivity of *Institution* with the private sector and academic institutions. This concept includes G2C, G2B, G2G, C2G and B2G, and indirectly to the B2B, C2C, C2B and B2C concepts.

Project phases / lots represent separate parts of the project, where each lot can be regarded as a project within a project, i.e. can be defined by any set of data from the parent project. The basic rule is that the data inside phases must always be a subset of the data of the main project.

Status of the project represents the status of the project during the reporting period, and is defined as *planned, in progress*, *completed* or *postponed.*

Start date represents the start date of the project.

The planned completion date of the project is defined as the expected end of the project.

The actual completion date is the date when the project is declared as completed. The difference between the actual and planned completion of the project is one of the indicators of the success of the project.

Sources of funding are defined by *Institution*(s), funds and partners who provide funding for the project and the clear indication of the share of these funds in the overall project.

The total amount of financing planned and **total amount of financing executed** are representing the difference between budgeted costs and actual costs and are also an indicator of the success of the project.

The submitted documentation represents all the documentation that came with the *ICT Project*, as evidence that all entries are correct.

Notes represent an optional field for the parameters of the project that cannot be displayed existing fields.

IV. CONCEPTUAL MODEL OF INFLUENCE OF PROJECTS

In order to enable determination of influence of projects in context of achieving the broader goals of electronic public administration, in addition to model of *ICT Project,* it is necessary to define model for evaluating the influence of specific *ICT Project* on electronic governance. This section describes proposed conceptual model – **Electronic Administration and Level of Information System (EALIS) matrix.**

EALIS matrix has three dimensions.

Dimension X represents the service(s) of public administration that are subject of the specific project.

The variable X takes the value from the set corresponding to a field *Public Administration Services* (a collection of 35 entries corresponding to legal framework and practice for acceding EU).

Dimension Y represents the level of service of public administration achieved through implementation of specific project and takes value from the set corresponding to a field *Degree of public administration services* (a collection of 5 entries corresponding to degrees of services).

Dimension Z is the time dimension, as the basis for monitoring the continuous development of electronic government.

One way to implement proposed conceptual model is to quantify dimensions X and Y adequately, and then to represent a single project as a function $f(x, y, z, a)$, where a is an input parameter that specifies the context of evaluation, and the value of the function is a "success" indicator, i.e. value of the impact of the project on the entire system.

Improvement of the entire system is measured by the sum of the results of all projects by time

$$Q = \sum_{z=t_1}^{t_2} f(x, y, z, a).$$

By changing the input parameter a it is possible to get a large number of indicators of the entire system.

With the proposed model it is possible to calculate the following indicators:

1. the overall level of development of electronic government for a certain period of time;
2. level of development of specific aspects of electronic governance indicators through time;

3. level of development of specific aspects of electronic government through specific project or projects.

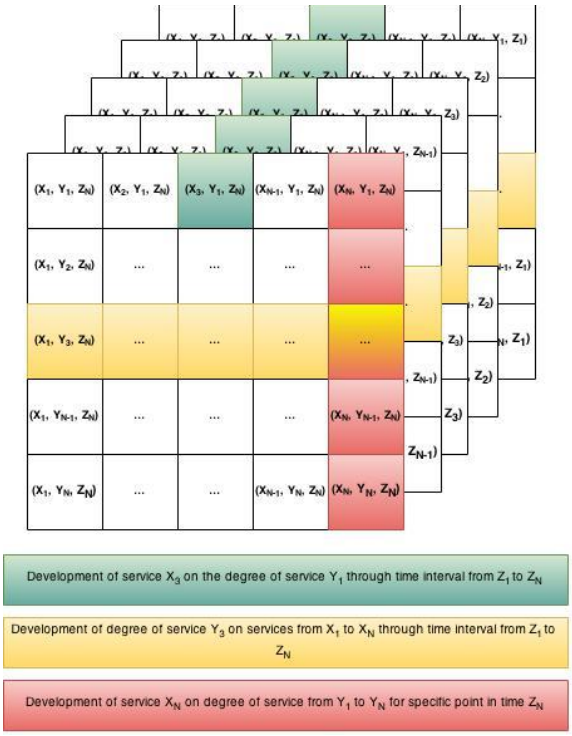


Figure 4.1 – Specific reports of EALIS matrix

Figure 4.1 represents an illustration of generating individual reports based on the EALIS matrix.

By applying mathematical operations on indicators of individual reports by some of the dimensions of the matrix aggregated reports can be obtained, and the diversity in reporting can be achieved by varying the input parameter a which can be used from corresponding metadata of *ICT Project* entity. One example of the input parameters a is the metadata of the project entity that can be quantified, and the form of functional dependencies is one of the central topics for further research.

In this way, it is possible to get EALIS matrix of successfully implemented projects or EALIS matrix of funding sources, etc. The main challenge of the proposed method of assessment is an adequate quantification of individual indicators and the definition of consistent metrics (for example, the metric for the degree of service).

V. CONCLUSION

Modelling of electronic government ICT projects proposed in this paper is aimed to provide indicators for assessing the impact of the project(s) on the broader goals of electronic public administration development. The practical aim is to provide indicators that could help to direct ICT projects planning and implementation towards achievements of the broad objectives of electronic government.

By combining current initiatives and methodologies for managing ICT projects, such as *„Database on ICT Projects in PA of Republic of Srpska – Methodology and Instructions for Forms Filling“* and *„Guidelines for providing expert opinion on ICT projects submitted to AISRS“*, this paper proposed methodology for contextual modelling of ICT projects and its application at the public administration of the Republic of Srpska.

Considering the context of ICT projects in public administration, the model identifies four entities. The *Institution* is an entity that is the main carrier of project activities, *ICT project* is the operating entity, while entities *Service* and *Citizen* represent base mechanism for evaluation of achievements in the development of electronic government. The semantics of ICT project within the context of achieving the broader goals of electronic governance is defined by semantics of metadata for the four identified entities. In addition, this paper proposes a new conceptual model of evaluation called EALIS matrix that includes the dimension of electronic service, degree of electronic service and time component for implementation.

It is planned that further research is going in two directions.

The first one will include further developing of model of metadata from the standpoint of the aspects of evaluation, as well as advancements of the evaluation model regarding allowed input data (imprecise, linguistic, etc.) and indicators calculation.

The second direction will include research related to machine-readable representation of the model, evaluation process automation, and presentation of the evaluation results.

REFERENCES

[1] Stevenson, A. „ *Oxford Dictionary of English* “, Oxford University Press, 2010.

[2] *„A Guide to the Project Management Body of Knowledge”,* 3rd Edition, Project Management Institute, Pennsylvania, USA, 2004.

[3] Alabau A. *“The European Union and its eGovernment development Policy”,* FUNDACION VODAFONE ESPAÑA, 2004.

[4] *„Guidelines for providing expert opinion on ICT projects submitted to AISRS “,* Agency for Information Society of Republic of Srpska, 2013. (in Serbian)

[5] *„Database on ICT Projects in PA of republic of Srpska – Methodology and Instructions for Forms Filling“,* Agency for Information Society of Republic od Srpska*, 2009.* (in Serbian)

[6] EU Acquis communautaire, http://ec.europa.eu/enlargement/policy/conditions-membership/chapters-of-the-acquis/index_en.htm, visite: 02.01.2015.

[7] Radinković M., Latinović M. „A Strategic Framework for eGovernment and eServices Development in the World“, Modern Government, Banja Luka, January 2010. (in Serbian)

Managing PhD promotions and register of doctors in CRIS UNS

Bojana Dimić Surla*, Lidija Ivanović**
* Faculty of Sciences, University of Novi Sad, Novi Sad, Serbia
** Faculty of Education, University of Novi Sad, Sombor, Serbia
e-mail: bdimic@uns.ac.rs, lidija.ivanovic@pef.uns.ac.rs

Abstract—Paper presents the module for managing and organizing PhD promotions at the University of Novi Sad. The module is integrated into CRIS UNS (Current Research Information System University of Novi Sad) that is based on international standards of representing research data. The module is implemented as web application and supports the process that begins with selecting candidates (from CRIS UNS system) for several forthcoming promotions, following by managing the lists of selected candidates, printing necessary information, and finishing with entering the promotion date, updating the list of all candidates promoted at the University for input into the official register.

I. INTRODUCTION

Promotion of the PhDs is a ceremonial act in which the Rector of the University promotes the candidates into doctors. For each promotion the officials at the Rectorate prepare a list of candidates that have defended dissertation in the certain time period before the scheduled promotion. The process of selecting candidates for promotion as well as updating the official register of the promoted doctors are automated and implemented within the web application of CRIS UNS.

CRIS (*Current Research Information System*) based on CERIF data model (Common European Research Information Format) is meant for processing all relevant entities of research activity and is considered very important for the development of science [1]. CERIF (Common European Research Information Format) is the standard that proposes the data model that allows the interoperability among research management systems [2]. The CRIS systems support entities that contain data about researchers, projects, scientific conferences, institutions, published results, etc. In the last couple of years, the most scientific institutions already have or are in the process of implementation of the CRIS system.

The system CRIS UNS [3,4] was developed for the needs of the University of Novi Sad according to the recommendations of non-profit organization euroCRIS (http://www.eurocris.org/). The implementation of the system started 2009 and it is publically available at http://www.cris.uns.ac.rs/. Two main requirements for the specification and implementation of the system were the compliance with the international standards in the field of representing scientific-research data on one side and fulfilling the local requirements specific to the University and Republic of Serbia within which the University was established. Moreover, the CRIS UNS is implemented as web application using Web 2.0 technologies for creation of user-friendly interface.

Among other types of scientific results, CRIS UNS supports the repository of dissertations defended at the University of Novi Sad that is integrated with the ETD-MS-compatible repositories as described in the papers [5,6,7]. Migration of data from the former dissertation repository DIGLIB UNS was described in the paper [8]. The ontology for presenting data about thesis and dissertation in the Semantic Web technology is given in [9]. Public service for searching repository of dissertations defended at the University of Novi Sad is available at:

http://www.cris.uns.ac.rs/searchDissertations.jsf.

The promotion of the dissertation and candidate is the very last step of the dissertation lifecycle that starts with the submission of the dissertation by the student. More details about this process are given in the next section. In the third section we give the use case diagram that describes the functional requirements of the module for organizing promotions. The data model that is behind the module and input parameter to official register is presented in fourth section. The implementation and screen forms are described in fifth section. The main purpose of implementation the module for managing dissertations and promotion is the implementation of the official registry of doctors that is described in sixth section. At the end, we provide some conclusions and summary of the work presented in the paper.

II. PROCESS OF COLLECTING DATA FOR PROMOTION

The procedure of submitting and defense of the dissertation consists of the well-defined steps and involves several entities, and these are Student service of the Faculty, the Faculty Council, the University Senate and the commission for evaluation and defense of the dissertation. For some steps in this process the data about the candidate and the dissertation are entered in CRIS UNS and in that way become available in the module for promotion.

After student finishes his/her dissertation, he needs to hand in the printed and electronic version of the dissertation to the Student service of the Faculty on which the dissertation is going to be defended. Student service then sends the request to the Faculty Council to establish the commission for evaluation and defense of the dissertation and the dissertation is handed to the commission members. After some time period, the

commission writes the evaluation report and submits it to the Student service at the Faculty.

By receiving the report, Student service enters some data about dissertation to the digital library within CRIS UNS including dissertation metadata such as author's name, the language of the dissertation, title of the dissertation and scientific field. The institution on which the dissertation is going to be defended is also stored in the CRIS UNS database and is obtained from the user account as the user form the Student service at the Faculty have to be logged in for entering these data. In addition to entering metadata, the Student service uploads the electronic version of the dissertation and the commission report for a public review. The documents are publically available in CRIS UNS web application for 30-day period.

After 30 days Student service sends the commission report to the Faculty Council together with the remarks and comments received during the public review. The Faculty Council then makes the decision about accepting or rejecting the report, or to require the review of the report. The Faculty Council sends the decision to the Student service. If the decision is accepting or rejecting (with the explanation) the commission report, then the Student service sends the report to the University Senate. If Senate gives the agreement on commission positive report about the quality of the dissertation, the student can defend dissertation, and the public defense is scheduled by the Student service.

If student successfully defend the dissertation, the Student service at the Faculty enters in CRIS UNS all necessary information for digital library and all data for promotion and the register.

In the described procedure, all metadata necessary for the organizing promotion become available in CRIS UNS and are further processed by the module that is the subject of this paper.

III. FUNCTIONAL REQUIREMENTS OF THE PROMOTION MODULE

Figure 1 shows the use case diagram that describes the functional requirements of the module for scheduling promotions.

The Rectorate official in charge of organizing promotions selects the candidates for promotion using the metadata entered by the Student services at the Faculties on which the dissertations were defended.

The system needs to support creating several lists for several forthcoming promotions. The lists are created by selecting candidates from the list of all candidates that have been sent to promotion by the Faculties. The lists are created over time, so the user can save the list and finishes it later.

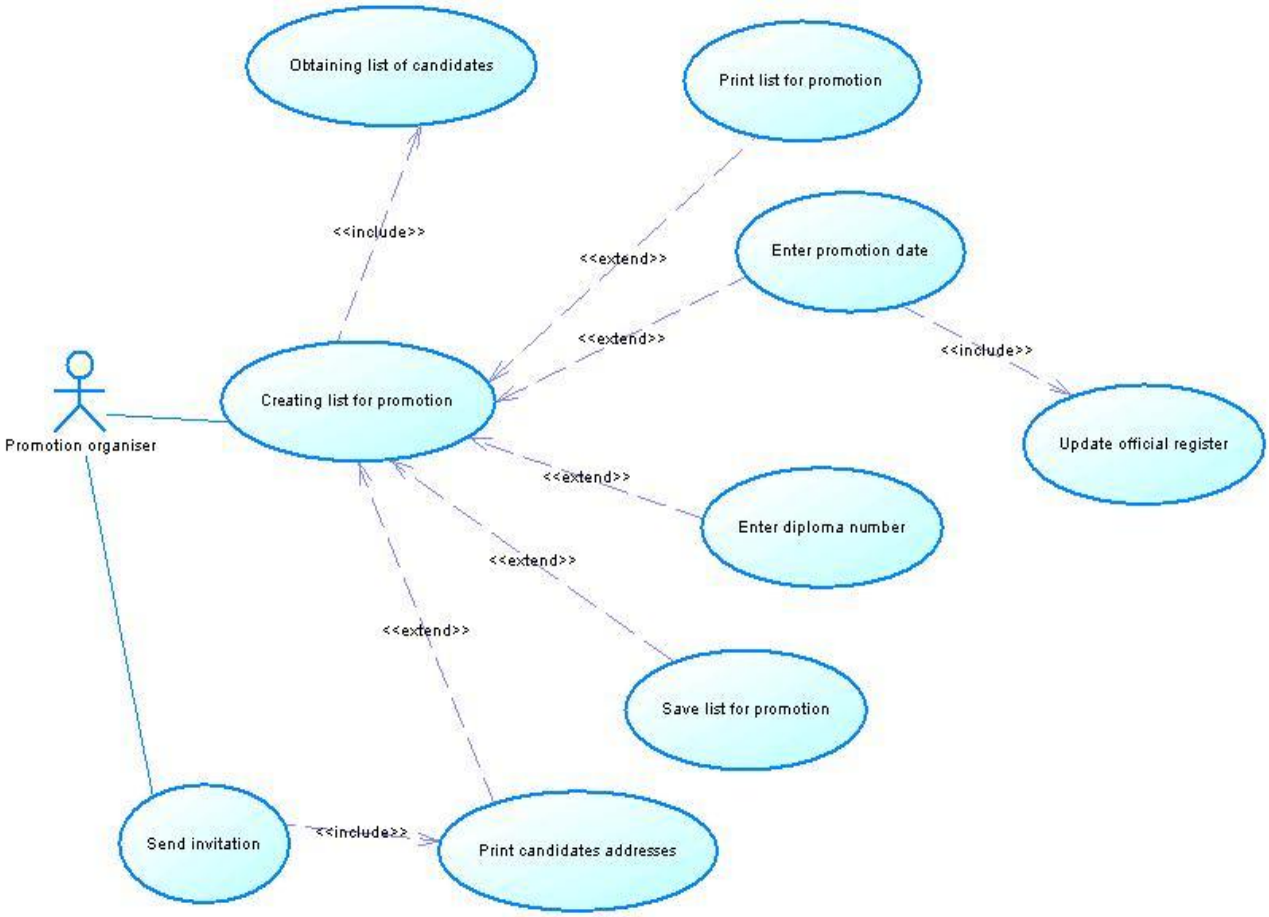


Figure 1. Use case diagram of the module for organizing promotion

The created lists for promotion are also used for gathering data for sending invitations and official records, so system needs to support printing the list for promotion and printing addresses.

For some candidates, and these are students that have defended PhD thesis by old program without finishing PhD studies by Bologna declaration program, diploma numbers are entered in this module. For other candidates, diploma numbers together with the supplement numbers are entered in CRIS by the Faculties.

After the promotion ceremony, the process is finished by entering the promotion date for the candidates that were promoted and updating the register of doctors.

IV. METADATA ABOUT DEFFENDED DISSERTATION – REGISTER ENTRY

For the implementation of the register as well as the module for organizing promotion at the University of Novi Sad the CRIS UNS data model [3] was extended by one more entity called RegisterEntry. This entity represents metadata about dissertation and author himself.

The data about author are his/her first and last name, mother's and father's name or guardian's name, date and place of birth.

The special properties are reserved for candidate's previous education such as the name and place of the previously graduated institution and name of author's previous academic degree (before PhD). The author's academic degree differs in case he has finished the old program (magister) or new program by Bologna declaration (master).

In addition to information about author, the RegisterEntry entity contains data about dissertation including title of the dissertation, name and place of the institution on which the dissertation was defended, advisor(s) and commission member, date on which dissertation was defended and academic degree obtained after dissertation defense. The special property is reserved for the CRIS UNS control number of the dissertation, identifier of the dissertation as a scientific result within CRIS. In case of a new program, i.e. PhD studies, the candidate gets the mark on public defense which is also stored in the RegisterEntry entity.

As was mention in previous section, the input of the diploma number also depends on the program. If candidate has finished PhD studies by the new program, the Student service on the Faculty enters the diploma number together with the supplement number. In case of old program, the official in the Rectorate in charge of promotion enters the diploma number. These pieces of information are stored in different properties of RegisterEntry.

Properties that are especially important for the promotion module are contact details of the candidate that are used for sending invitations, and these are street and number, place, postal code, city, country, email and phone number.

The system supports managing several forthcoming promotions and one property is reserved for the name of the future promotion. The property is defined as arbitrary string and user usually enter the date or the month of the future promotion.

After the promotion, the promotion date is stored in the related property.

Some properties are especially important for updating and generating register, and these are:

REGISTERENTRYID – ordinal number of the dissertation in the register,

ACADEMICYEAR – academic year in which the dissertation was defended,

ACADEMICYEARNUMBER - ordinal number of the dissertation within the academic year.

V. IMPLEMENTATION AND SCREEN FORMS

CRIS UNS system is implemented as web application using Web 2.0 technology. Other technologies used for the implementation are Java for the business logic and implementation of data model, Apache Tomcat as web server and MySQL as database management system.

After logging in at the web application the officials at the Rectorate in charge of scheduling PhD promotions open the web page for selecting candidates for promotion, shown in figure 2.

The functionality is organized in two tabs. The first tab (Candidates) provides the list of all candidates for promotion that defended dissertation and are not yet promoted. All candidates are presented in a table, with basic data and one extra text field in every row. That text field is reserved for entering the name of the future promotion. The second tab (Prepare for promotion) that is shown on figure 2 contains all candidates selected in the first tab and grouped by the promotion name. The user chooses the promotion name from the drop down menu and gets all candidates that are linked to that promotion. All data on figure 2 are given in Serbian (Cyrillic) like they are stored in the database and which is in accordance with the rules proposed by the law.

For the candidates that defended dissertation by old program it is enabled to enter the diploma number, and for those who finished the PhD studies the diploma and supplement numbers are entered by the Faculties and their text fields are disabled in the promotion module.

There are three actins that can be performed on the created list, the list can be printed, saved or added to the register. The action Print list opens the popup window in which user can choose different formats of printing the chosen list, such as the list of addresses for all candidates for sending invitation and list of candidates in special format used for publishing the official announcement for the promotion. By calling action Save list user stores all current lists for promotion in database and can finish editing them later. After user that is registered to be in charge for promotion logs in at the CRIS UNS he/she gets all previously prepared lists for promotion. Action

Add to register is called after the promotion. User enters the exact date of promotion after which the ids for the register entries i.e. the ordinal numbers for all dissertations promoted are generated and they are stored in the database. As the final step, the register can be generated, as described in the next section.

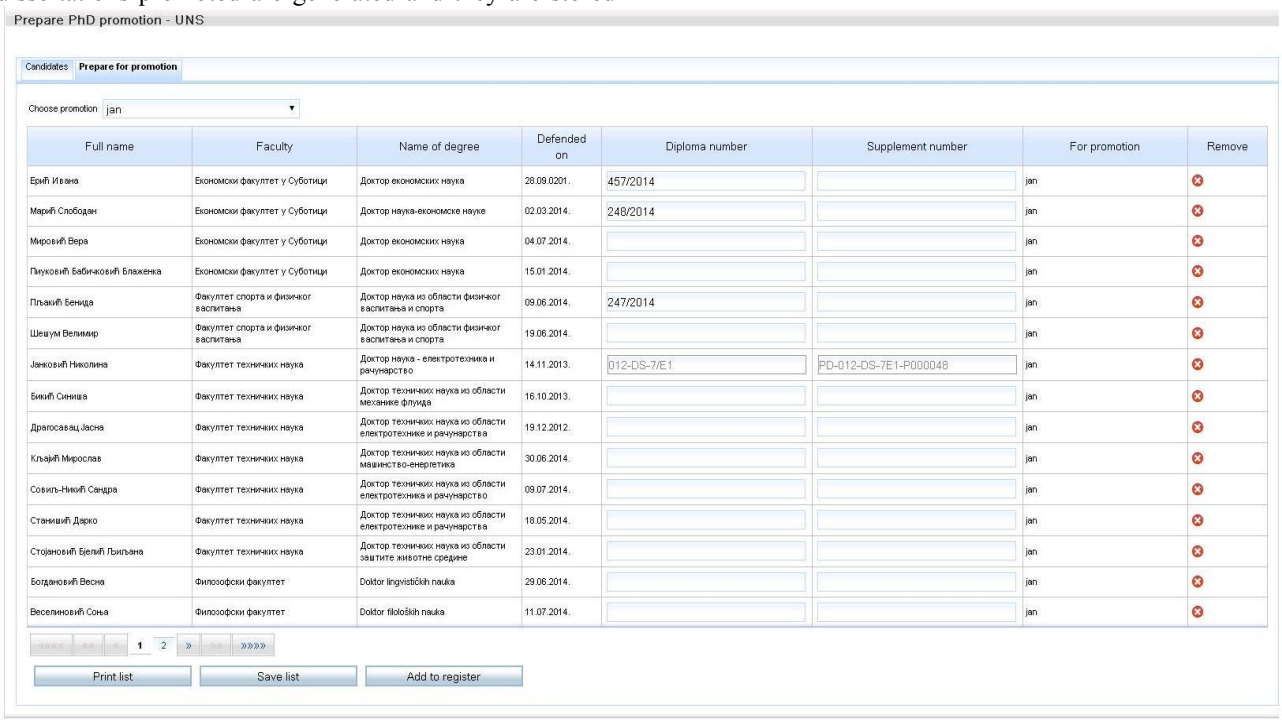


Figure 2. Screen form for organizing PhD promotion in CRIS UNS

VI. REGISTER OF DOCTORS

Every University in Serbia is obliged to keep the official register of all doctors that defended dissertation at that University. In case of University of Novi Sad, that register was kept manually, by handwriting all data proposed by the low.

By introducing CRIS UNS, and keeping records about dissertations and promotions in the information system, the register is created as a report in PDF format (Figure 3) by sending request to web application.

The official register is generated in CRIS UNS by calling Java Servlet on specific URL. Servlet supports several cases of creating register in real time, one of which is generating the complete register by including all dissertations stored in CRIS UNS that has promotion date and ordinal number. The second case is generating register for the specific period by assigning start and end date. And the last is generating register by including dissertation from the specific institution. This case is used in purpose of validation and checking the quality of the entered data.

The official register is implemented as a report using the open source report engine - Jasper Report Library (https://community.jaspersoft.com/project/jasperreports-library). The input in the report engine is the list of RegisterEntry entities that are described in section 4.

As shown in Figure 3, the report is given in a form of a table, and every dissertation is presented with all data from the RegisterEntry entity. The register entries are grouped by the academic year, and are sorted by the ordinal number of the dissertation. The labels and data in register shown on figure 3 are given in Serbian (Cyrillic) as this is the official document and the language, labels and the format are proposed by the low.

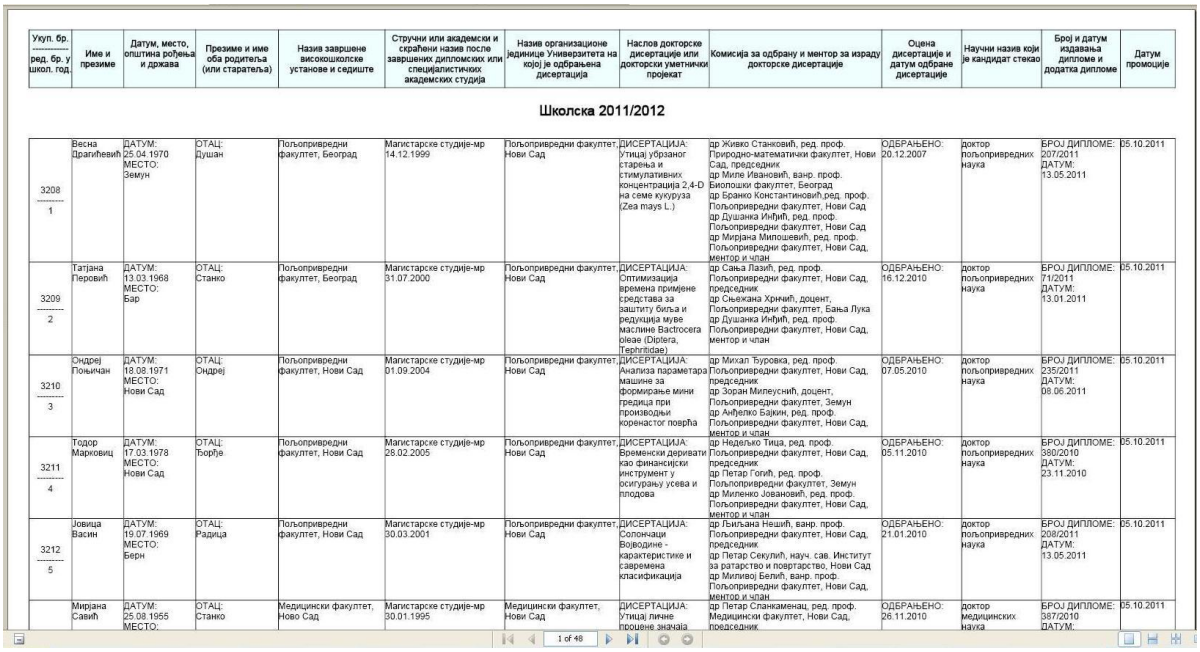
VII. CONCLUSION

The main advantage of having the application described in this paper is having the complete lifecycle of the dissertation managed in the web application. The application is accessed by different users that have the specific role in the process.

Moreover, in this case we integrated the process of organizing promotions and updating the register with the digital library of dissertations and information system for storing and managing data about scientific research (CRIS).

The very last step in processing the dissertation is promoting the candidate into doctors which is done in the ceremonial promotion organized by the Rectorate. In this paper we described how we used data collected in CRIS UNS in module for managing the organization of that promotion.

The main purpose of developing module for organizing promotions in CRIS is having the environment for automatic generation of official register. The layout of the register, as well as the set of data included in the register are proposed by the low and now is implemented as a report and integrated into web application.



Укуп. бр. / ред. бр. у школ. год	Име и презиме	Датум, место, општина рођења и држава	Презиме и име оба родитеља (или старатеља)	Назив завршене високошколске установе и седиште	Стручни или академски и скраћени назив после завршених дипломских или специјалистичких академских студија	Назив организационе јединице Универзитета на којој је одбрањена дисертација	Наслов докторске дисертације или докторски уметнички пројекат	Комисија за одбрану и ментор за израду докторске дисертације	Оцена дисертације и датум одбране дисертације	Научни назив који је кандидат стекао	Број и датум издавања дипломе и додатка дипломе	Датум промоције
Школска 2011/2012												
3208 / 1	Весна Драгићевић	ДАТУМ: 25.04.1970 МЕСТО: Земун	ОТАЦ: Душан	Пољопривредни факултет, Београд	Магистарске студије-мр 14.12.1999	Пољопривредни факултет, Нови Сад	ДИСЕРТАЦИЈА: Утицај убрзаног старења и стимулативних концентрација 2,4-D на семе кукуруза (Zea mays L.)	др Живко Станковић, ред. проф. Природно-математички факултет, Нови Сад, председник др Миле Ивановић, ванр. проф. Биолошки факултет, Београд др Бранко Константиновић,ред. проф. Пољопривредни факултет, Нови Сад др Душанка Ињић, ред. проф. Пољопривредни факултет, Нови Сад др Мирјана Милошевић, ред. проф. Пољопривредни факултет, Нови Сад, ментор и члан	ОДБРАЊЕНО: 20.12.2007	доктор пољопривредних наука	БРОЈ ДИПЛОМЕ: 207/2011 ДАТУМ: 13.05.2011	05.10.2011
3209 / 2	Татјана Перовић	ДАТУМ: 13.03.1968 МЕСТО: Бар	ОТАЦ: Станко	Пољопривредни факултет, Београд	Магистарске студије-мр 31.07.2000	Пољопривредни факултет, Нови Сад	ДИСЕРТАЦИЈА: Оптимизација времена примене средстава за заштиту биља и редукција муве маслине Bactrocera oleae (Diptera, Tephritidae)	др Сања Лазић, ред. проф. Пољопривредни факултет, Нови Сад, председник др Снежана Хрнчић, доцент, Пољопривредни факултет, Бања Лука др Душанка Ињић, ред. проф. Пољопривредни факултет, Нови Сад, ментор и члан	ОДБРАЊЕНО: 16.12.2010	доктор пољопривредних наука	БРОЈ ДИПЛОМЕ: 71/2011 ДАТУМ: 13.01.2011	05.10.2011
3210 / 3	Ондреј Пољичан	ДАТУМ: 18.08.1971 МЕСТО: Нови Сад	ОТАЦ: Ондреј	Пољопривредни факултет, Нови Сад	Магистарске студије-мр 01.09.2004	Пољопривредни факултет, Нови Сад	ДИСЕРТАЦИЈА: Анализа параметара машине за формирање мини гредица при производњи коренастог поврћа	др Михал Ђуровка, ред. проф. Пољопривредни факултет, Нови Сад, председник др Зоран Милеуснић, доцент, Пољопривредни факултет, Земун др Анђелко Бајкин, ред. проф. Пољопривредни факултет, Нови Сад, ментор и члан	ОДБРАЊЕНО: 07.05.2010	доктор пољопривредних наука	БРОЈ ДИПЛОМЕ: 235/2011 ДАТУМ: 08.06.2011	05.10.2011
3211 / 4	Тодор Марковић	ДАТУМ: 17.03.1978 МЕСТО: Нови Сад	ОТАЦ: Ђорђе	Пољопривредни факултет, Нови Сад	Магистарске студије-мр 28.02.2005	Пољопривредни факултет, Нови Сад	ДИСЕРТАЦИЈА: Временски деривати као финансијски инструмент у осигурању усева и плодова	др Неделько Тица, ред. проф. Пољопривредни факултет, Нови Сад, председник др Петар Гогић, ред. проф. Пољопривредни факултет, Земун др Миленко Јовановић, ред. проф. Пољопривредни факултет, Нови Сад, ментор и члан	ОДБРАЊЕНО: 05.11.2010	доктор пољопривредних наука	БРОЈ ДИПЛОМЕ: 380/2010 ДАТУМ: 23.11.2010	05.10.2011
3212 / 5	Јовица Васин	ДАТУМ: 19.07.1969 МЕСТО: Берн	ОТАЦ: Радица	Пољопривредни факултет, Нови Сад	Магистарске студије-мр 30.03.2001	Пољопривредни факултет, Нови Сад	ДИСЕРТАЦИЈА: Солончаци Војводине - карактеристике и савремена класификација	др Гњиљана Нешић, ванр. проф. Пољопривредни факултет, Нови Сад, председник др Петар Секулић, науч. сав. Институт за ратарство и повртарство, Нови Сад др Миливој Белић, ванр. проф. Пољопривредни факултет, Нови Сад, ментор и члан	ОДБРАЊЕНО: 21.01.2010	доктор пољопривредних наука	БРОЈ ДИПЛОМЕ: 208/2011 ДАТУМ: 13.05.2011	05.10.2011
	Мирјана Савић	ДАТУМ: 25.08.1955 МЕСТО:	ОТАЦ: Станко	Медицински факултет, Ново Сад	Магистарске студије-мр 30.01.1995	Медицински факултет, Нови Сад	ДИСЕРТАЦИЈА: Утицај личне процене значаја	др Петар Сланкаменац, ред. проф. Медицински факултет, Нови Сад, председник	ОДБРАЊЕНО: 26.11.2010	доктор медицинских наука	БРОЈ ДИПЛОМЕ: 387/2010 ДАТУМ:	05.10.2011

Figure 3. Official register of the doctors at the University of Novi Sad

ACKNOWLEDGMENT

The work is supported by the Ministry of Education and Science of the Republic of Serbia, through Project No. OI174023: "Intelligent techniques and their integration into wide-spectrum decision support".

REFERENCES

[1] Zimmerman, E., "CRIS-Cross: Current Research Information Systems at a Crossroads," *Proceedings of the 6th International Conference on Current Research Information Systems*, University of Kassel, August 29 - 31, 2002, pp. 11-20

[2] Asserson, A., Jeffery, K. and Lopatenko, A., "CERIF: Past, Present and Future: An Overview," *6th International Conference on Current Research Information Systems,* University of Kassel, 2002, August 29 - 31

[3] Ivanović, D., Surla, D. and Konjović, Z., "CERIF compatible data model based on MARC 21 format," *The Electronic Library*, 2011, vol. 29, no. 1, pp. 52-70

[4] Ivanović, D. et al, "A CERIF-compatible research management system based on the MARC 21 format," *Program: Electronic Library and Information Systems*, 2010, vol. 44, no. 1, pp. 229-251

[5] Ivanović, L., Ivanović, D. and Surla, D., "A data model of theses and dissertations compatible with CERIF, Dublin Core and EDT-MS," *Online Information Review*, 2012, vol. 36, no. 4, pp. 568-586

[6] Ivanović, L., Ivanović, D., Surla, D., –Integration of a Research Management System and an OAI-PMH Compatible ETDs Repository at the University of Novi Sad, Republic of Serbia", *Library Resources & Technical Services*, 2012, vol. 56, vo. 2, pp. 104-112

[7] Ivanovic, L., Ivanovic, D., Surla, D., & Konjovic, Z., "User interface of web application for searching PhD dissertations of the University of Novi Sad," In *Intelligent Systems and Informatics (SISY)*, 2013 IEEE 11th International Symposium on (pp. 117-122)

[8] Ivanovića, L. and Surla, D., –A software module for import of theses and dissertations to CRISs", *Proceedings of the CRIS 2012 Conference*, Prague, June 6-9. 2012, pp. 313-322

[9] Ivanović, L. et al., "CRISUNS ontology for theses and dissertations," In:*Proceedings of the 2nd International Conference on Information Society Technology - ICIST 2012* (CD), Kopaonik, 2012

Evaluation of the implementation of the "eAdministration Strategy of Provincial Authorities"

Milan Paroški*, Vesna Popović*, Dušan Surla**, Zora Konjović***
* Government of the AP of Vojvodina/Office for Joint Affairs of Provincial Bodies, Novi Sad, Republic of Serbia
**Faculty of Sciences/Department of Mathematics and Informatics, Novi Sad, Republic of Serbia
*** Faculty of Technical Sciences/Department of Computing and Control, Novi Sad, Republic of Serbia
milan.paroski@vojvodina.gov.rs; vesna.popovic@vojvodina.gov.rs; surla@uns.ac.rs; ftnzora@uns.ac.rs;

***Abstract*— Within the framework of methodology adopted for drawing up the document "Strategy of eAdministration of Provincial Authorities with the Action Plan until 2015", the study was conducted with the aim to evaluate previous efforts and the adoption of the services, applications and infrastructure objects completed in the examined period (2007-2013). The purpose of this paper is to present several aspects related to defined methodology, as well as a part of the basic results of the study on evaluation of the eAdministration development at provincial level until 2013.**

I. INTRODUCTION

Pursuant to the Decision on the Provincial Administration Reform and Development ("Official Journal of the APV", No. 14/2006), in September 2007, the Executive Council of the Autonomous Province of Vojvodina adopted the Decision on the eAdministration Strategy of Provincial Authorities ("Official Journal of the APV", No. 18/2007).

The Government of the AP Vojvodina has engaged IT experts from the University of Novi Sad to prepare a document by means of which they would establish how the aforementioned Strategy was implemented by the beginning of 2013 and draw up a revision, for the purpose of continuing development of the eAdministration of Provincial Authorities according to the revised Strategy in the five-year period from 2013 to 2018.

The following methodology for drawing up the new strategic document has been applied:

1. During the drawing up of the document, a framework, goals and principles specified in the initial document titled "eAdministration Strategy of Provincial Authorities" (2007) were adopted.
2. As a basis for planning, data from two types of documents were used: planning documents and situation evaluation documents.
3. During the drawing up of the document, the Provincial Administration reorganization in the period from 2010 to 2012 was taken into account.
4. Methodology for managing continuous development of the eAdministration of Provincial Authorities was defined.

Evaluation should start together with the eAdministration initiatives, from the early conceptualization and design phase onwards. Evaluation must compare actual intervention results with original intentions. Within the framework of preparation of this strategic document, the study was conducted with the aim to evaluate the previous implementation of the "eAdministration Strategy of Provincial Authorities" which was adopted in 2007. The recommendations from available literature on acceptance and adoption of eGovernment by its intended users were used in study preparation. [1], [2].

As the result of activities of the Working Group, the Decision on the "Strategy of eAdministration of Provincial Authorities with the Action Plan until 2015" was enacted in July 2013 ("Official Journal of the APV", No. 26/2013).

II. SURVEY RESEARCH

Several aspects related to the evaluation of the situation of eAdministration implementation in Provincial Authorities were examined in the survey. Working Group for the development of the document "Strategy of eAdministration of Provincial Authorities with the Action Plan until 2015" prepared and administered a questionnaire to participants by mail. The participants were chosen as a representative sample of IT experts, authorized employees of the provincial administrative authorities. It is important to note that this survey is based on a closed-invitation-only sample in each participating Provincial Authority. The completed questionnaires were returned by 17 provincial bodies:

1. Assembly of the AP of Vojvodina
2. Provincial Government
3. Provincial Secretariat for Economy
4. Provincial Secretariat for Culture and Public Information
5. Provincial Secretariat for Health Care, Social Policy and Demography
6. Provincial Secretariat for Education, Administration and National Communities
7. Provincial Secretariat for Interregional Cooperation and Local Self-Government
8. Provincial Secretariat for Science and Technological Development
9. Provincial Secretariat for Energy and Mineral Resources
10. Provincial Secretariat for Urban Planning, Construction and Environmental Protection
11. Provincial Secretariat for Sports and Youth

12. Provincial Secretariat for Labour, Employment and Gender Equality
13. Office for Joint Affairs of Provincial Bodies
14. Professional Service for Implementation of the Integrated Regional Development Plan of the APV
15. Human Resources Management Service
16. Provincial Ombudsman
17. Directorate of Commodity Reserves of the APV

The questions in the questionnaire are divided into several groups. The first group of questions covers issues related to the document "eAdministration Strategy of Provincial Authorities" that was enacted in 2007. The answers were predefined. The participants were asked to answer the following questions:

- Had employees at your administrative authority been informed about the document "eAdministration Strategy of Provincial Authorities" before you received it enclosed with this questionnaire?
- Which of the employees at your authority/ body had read the document "eAdministration Strategy of Provincial Authorities" before you received it enclosed with this questionnaire?

The second group of questions relates to the projects which were implemented in the period 2007 – 2013 according to Action Plan of the document "eAdministration Strategy of Provincial Authorities". The participants were asked:

- Please evaluate to what extent the projects proposed in the document "eAdministration Strategy of Provincial Authorities" meet the overall needs of all provincial administrative authorities/bodies.
- Please evaluate to what extent the projects proposed in the document "eAdministration Strategy of Provincial Authorities" meet the needs of Your provincial administrative authority/body.

The third group of questions is related to the frequency of use and significance of services, applications and infrastructure objects completed in the defined period. The participants were asked to assign independently marks from 0 to 5 to assess an impact of each implemented service. Furthermore, the participants were asked to evaluate to what extent implemented systems meet the needs of their provincial administrative authority/body.

Finally, the participants were asked to evaluate the importance of planned projects which were not realized in the period 2007 – 2013.

A. *Awareness of the strategic document amongst employees*

Firstly, the level of awareness of employees about the existence and content of the document "eAdministration Strategy of Provincial Authorities" was investigated. The results are presented in Fig.1 and Fig. 2.

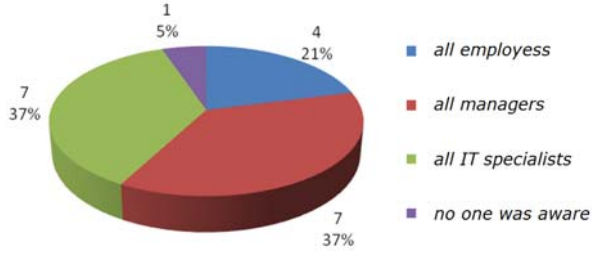


Figure 1. Previous awareness of employees about the existence of the document "eAdministration Strategy of Provincial Authorities"

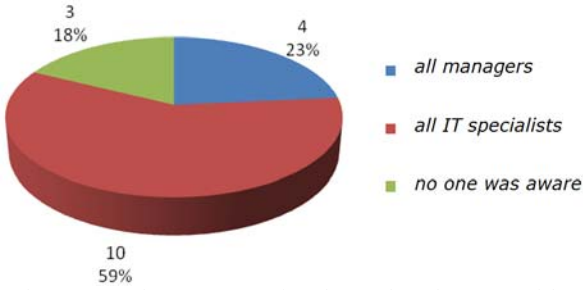


Figure 2. Previous awareness of employees about the content of the document "eAdministration Strategy of Provincial Authorities"

The study results are worrying. Even though the document was primarily focused towards the development of the IT support intended for the activities of the Provincial Administration, the content of the document before the conduction of the study was familiar mostly to those employed in ICT. Content of the document was familiar to the managers of only 23% provincial bodies who have participated in the study (Fig. 2). Also, response from 18% of administrative authorities covered by the study (Fig. 2), was that no one was aware of the content of the document.

Later on, the issue on how the projects proposed by the "eAdministration Strategy of Provincial Authorities" cover the needs of all authorities of the Provincial Administration, and in what extent proposed projects cover the needs of each individual authority of Provincial Administration, was examined. The obtained results are shown in Fig. 3. and Fig. 4. and they indicate that the projects cover only the basic needs (concerning over 80% of the provincial authorities).

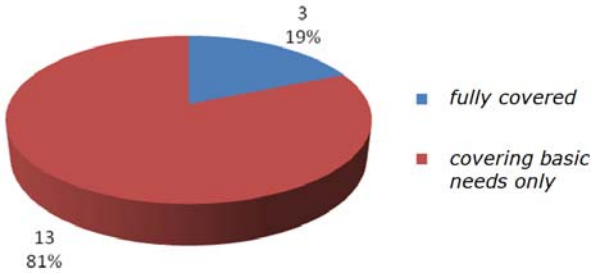


Figure 3. The estimate on the covering of needs of all authorities/bodies by projects proposed by the "eAdministration Strategy of Provincial Authorities"

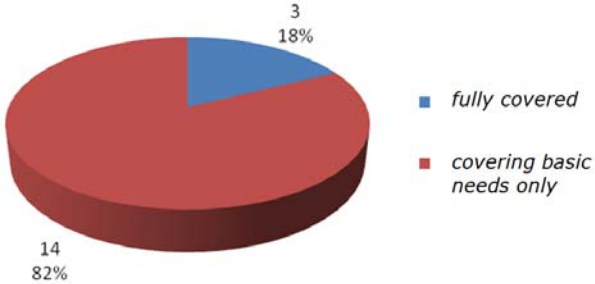


Figure 4. The estimate on the covering of the needs of individual provincial administration authorities/bodies by projects proposed by the "eAdministration Strategy of Provincial Authorities"

B. *Evaluation of eAdministration projects*

The second group of questions related to the estimation on the coverage of needs of the provincial authorities by the existing (already implemented) basic applications and infrastructure systems. The following applications, which were already in use, were selected as basic:

- *eDocumentus* system;
- *eRecordsManagementOffice* system;
- *eSessions* system [3];
- System for the *Inspection of the normative and legal acts of the Republic of Serbia*;
- System for the *Legal regulations and case law* (Ministry of Justice of the Republic of Serbia);
- System for the *Support for the organisation and conduction of exams*;
- *eLearning* system service (providing information and study material presentations);
- System for the *publishing of general information on the AP Vojvodina website*;
- *BISIS library service* system [4-9];
- *Online public procurement documentation* system;
- System for *Online information regarding subsidies and grants*;
- *eCompetitions* system [10];
- *Human Resources records* system;
- *Attendancy records (ID cards*) system;
- *System for the management of the use of the AP Vojvodina motor pool;*
- *Info-kiosk* system;
- *BISTreasury* integrated system for the preparation and performance of the budget and payroll accounting in provincial authorities.

The following infrastructure systems were selected [11]:

- Local computer network system;
- Internet access system;
- Video surveilance system;
- Electronic mail system;
- SMS system;
- Activities regarding the ECDL employees training programme.

The obtained results are presented in Fig. 5. and Fig. 6.

The frequency of the use of services, applications and infrastructure objects completed in the defined period was examined. The results are presented in Table I. The results show that the two most frequently used applications are *eDocumentus* and *BISTreasury*.

TABLE I.
The frequency of the use of applications and infrastructure objects completed in the examined period 2007-2013

System	N/A	Not used	Less than once a month	At least once a month	At least once a week	Every day
eDocumentus	0 0.00%	0 0.00%	0 0.00%	4 23.53%	3 17.65%	10 58.82%
eRecordsManagementOffice	4 23.53%	7 41.18%	1 5.88%	2 11.76%	1 5.88%	2 11.76%
eSessions	4 23.53%	5 29.41%	2 11.76%	1 5.88%	3 17.65%	2 11.76%
Inspection of normative and legal acts of the Republic of Serbia	3 17.65%	3 17.65%	0 0.00%	6 35.29%	3 17.65%	2 11.76%
Legal regulations and case law (Ministry of Justice of the Republic of Serbia)	4 23.53%	5 29.41%	0 0.00%	4 23.53%	2 11.76%	2 11.76%
Support for the organisation and conduction of exams	5 29.41%	8 47.06%	3 17.65%	0 0.00%	0 0.00%	0 0.00%
eLearning system service (providing information and study material presentations)	6 35.29%	5 29.41%	4 23.53%	1 5.88%	0 0.00%	1 5.88%
Publishing general information on the AP Vojvodina website	6 35.29%	3 17.65%	3 17.65%	1 5.88%	1 5.88%	3 17.65%
BISIS library service	5 29.41%	1 5.88%	5 29.41%	4 23.53%	0 0.00%	2 11.76%
Online public procurement documentation	4 23.53%	7 41.18%	1 5.88%	3 17.65%	1 5.88%	1 5.88%
Online information regarding subsidies and grants	6 35.29%	8 47.06%	1 5.88%	1 5.88%	1 5.88%	0 0.00%
eCompetitions	5 29.41%	8 47.06%	0 0.00%	2 11.76%	0 0.00%	2 11.76%
Human resources records	2 11.76%	3 17.65%	5 29.41%	2 11.76%	2 11.76%	3 17.65%
Attendancy records (ID cards)	2 11.76%	4 23.53%	1 5.88%	4 23.53%	3 17.65%	3 17.65%
Management on the use of the AP Vojvodina motor pool	5 29.41%	8 47.06%	1 5.88%	0 0.00%	0 0.00%	3 17.65%
Info-kiosk	8 47.06%	7 41.18%	1 5.88%	1 5.88%	0 0.00%	0 0.00%
BISTreasury integrated system for the planning and managing of the budget and payroll accounting in provincial authorities	1 5.88%	0 0.00%	1 5.88%	1 5.88%	3 17.65%	11 64.71%

Based on these surveys, as shown at Fig. 5. and Fig. 6, it can be concluded that the needs are better covered by infrastructure systems (there were no interviewees who feel that not even the most basic needs are covered and 47% of the interviewees consider that their needs were fully covered), while the coverage of the basic needs by applications is not satisfactory (only 23% of the interviewees feel that their needs are fully covered, while 6% consider the applications do not cover even the basic needs).

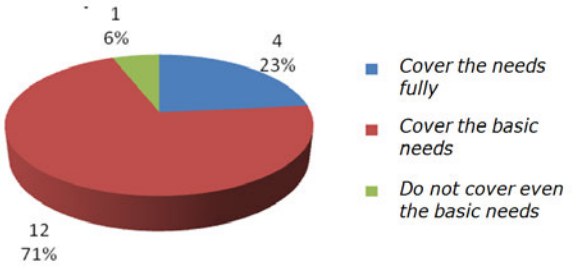


Figure 5. Coverage of the needs of provincial authorities by implemented basic applications

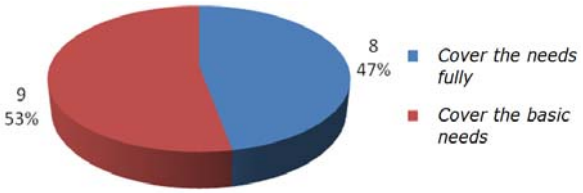


Figure 6. Coverage of the needs of provincial authorities by implemented infrastructure systems

In addition to the aforementioned, the importance of unrealized projects for the Provincial Administration authorities was examined. The possible answers were predefined. The interviewees were asked to rate the importance of the following unrealized projects in their everyday work:

- Introduction of the *eDocumentus* system in the Assembly of AP Vojvodina;
- *eArchive* system – recording and filing of cases;
- *eAdministrativeProcedeengs* system – Conducting administrative procedures and decisions regarding administrative affairs;
- *eAPVAdministrative practice system* – Administrative practice of the provincial authorities;
- *eCharging system* – collection of provincial taxes and service charges;
- *eDecisions* system – Monitoring the delivery and execution of decisions;
- Integrated system of provincial administrative registers;
- *eAPVEnrollement* system – enrollment into the higher education institutions in AP Vojvodina;
- *eCultureAPV system* – information system for affairs in the field of culture, within the competencies of AP Vojvodina;
- *eHealthAPV system* – information system for for affairs in the field of healthcare, within the competencies of AP Vojvodina;
- *eSocialProtectionAPV system* – information system for for affairs in the field of social policy, within the competencies of AP Vojvodina;
- *ePublicProcurements system* – 4th level service for the support in conducting public procurement procedures in provincial authorities;
- *Infrastructure of the spatial data of AP Vojvodina* – GIS system for spatial data in AP Vojvodina;
- *eEUCompetitions system* – monitoring the competitions of the EU;
- *Call centre* for provincial authorities;
- *eOrdering, eSituationMonitoring and eReporting systems* – storage facilities: distribution and archiving of materials, equipment and spare parts;
- *eTechnicalSystem system*– maintenance and security of the provincial authorities' buildings;
- *ePublishing system* – preparation and publication of printed materials.

Among them, several projects were identified as critical, still not introduced systems. As 'critical unintroduced systems' were denoted systems that more than 25% of the provincial authorities have identified as critical or very important for their work. The results are shown in Fig. 7.

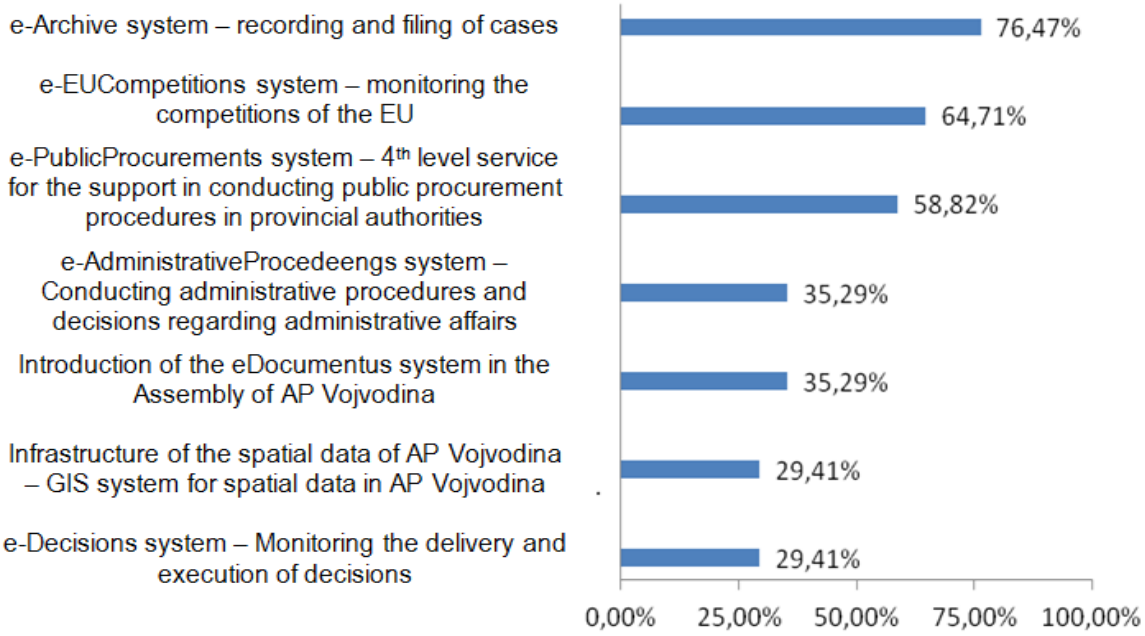


Figure 7. Critical unintroduced systems

The *eArchive, eEUCompetitions* and *ePublicProcurements* systems were identified as crucial for the activities of provincial authorities even though they were not introduced.

The results of the assessment of significance of all planned, but not introduced systems are shown in Table II.

Indicative is a high percentage of participants which are not aware of the importance of sectorial infrastructure systems for citizens and legal entities (*eHealthAPV, eSocialProtectionAPV, eCultureAPV*), as well as the stance that the implementation of such systems is not vital to the Provincial Authorities.

TABLE II.
Estimation of the significance of unintroduced systems for the work of provincial authorities

System	Significance assesment					
	Unknown	No significance	Little	Medium	Great	Crucial
eAPV Health Care system – health information system within the scope of competences of the AP Vojvodina	7 41.18%	6 35.29%	1 5.88%	2 11.76%	1 5.88%	0 0.00%
eAPV Social Protection system – social policy information system within the scope of competences of the AP Vojvodina	7 41.18%	6 35.29%	2 11.76%	1 5.88%	1 5.88%	0 0.00%
eCharging– collection of provincial taxes and service charges	7 41.18%	3 17.65%	3 17.65%	2 11.76%	2 11.76%	0 0.00%
eAPV Enrollment system – enrollment at higher education institutions in the AP Vojvodina	8 47.06%	2 11.76%	3 17.65%	2 11.76%	0 0.00%	2 11.76%
eAPV Culture system – culture information system within the scope of competences of the AP Vojvodina	7 41.18%	4 23.53%	2 11.76%	2 11.76%	2 11.76%	0 0.00%
eOrdering, eStockMonitoring and eReporting– storage facilities: distribution and archiving of materials, equipment and spare parts	7 41.18%	2 11.76%	3 17.65%	3 17.65%	1 5.88%	1 5.88%
eTechnical System – maintenance and security of the facilities of the Provincial authorities	9 52.94%	2 11.76%	3 17.65%	1 5.88%	1 5.88%	1 5.88%
Integrated system of the Provincial administrative registers	9 52.94%	1 5.88%	1 5.88%	3 17.65%	3 17.65%	0 0.00%
ePublishing– preparation and publication of printed materials.	6 35.29%	3 17.65%	1 5.88%	4 23.53%	2 11.76%	1 5.88%
eAPV Administrative Practice system– administrative practice of the Provincial authorities	4 23.53%	1 5.88%	3 17.65%	5 29.41%	4 23.53%	0 0.00%
Call centre for provincial authorities	5 29.41%	1 5.88%	3 17.65%	4 23.53%	2 11.76%	2 11.76%
eDecisions system – monitoring the delivery and execution of decisions	7 41.18%	1 5.88%	0 0.00%	4 23.53%	4 23.53%	1 5.88%
Infrastructure of the spatial data of AP Vojvodina – GIS system for spatial data in AP Vojvodina	4 23.53%	5 29.41%	2 11.76%	1 5.88%	5 29.41%	0 0.00%
Introduction of the eDocumentus system in the Assembly of AP Vojvodina	2 11.76%	4 23.53%	2 11.76%	3 17.65%	5 29.41%	1 5.88%
eAdministrativeProcedeengs– Conducting administrative procedures and decisions regarding administrative issues	5 29.41%	3 17.65%	1 5.88%	2 11.76%	6 35.29%	0 0.00%
ePublicProcurements system – 4th level service for the support in conducting public procurement procedures in provincial authorities	3 17.65%	0 0.00%	1 5.88%	3 17.65%	8 47.06%	2 11.76%
eEU Competitions system – monitoring the EU competitions for allocation of funds	1 5.88%	2 11.76%	1 5.88%	2 11.76%	10 58.82%	1 5.88%
eArchive system – recording and filing of cases	2 11.76%	0 0.00%	0 0.00%	2 11.76%	11 64.71%	2 11.76%

III. RESULTS AND DISCUSSION

The "eAdministration Strategy of Provincial Authorities" adopted by the Executive Council of the Autonomous Province of Vojvodina in 2007 is document in accordance with republic and European standards and values. It represents basis for government modernization and increase of efficiency of provincial administration. The integrated application of the ICT in provincial authorities is to enhance: efficiency, effectiveness, transparency, responsibility and economy in the work of Provincial Authorities and provincial civil servants. Simply put, the main aims are to improve the quality and availability of information and services provided to users by provincial civil servants.

The process of eAdministration introduction involves measuring actual progress and comparing it to planned, according to the adopted Action plan.

Data on the evaluation of implemented eAdministration projects were obtained by specified-purpose surveys. Results show that top management should provide more evidence of its commitment to the development and implementation of the eAdministration projects. Awareness of employees about the existence and the content of the document "eAdministration Strategy of Provincial Authorities" was not satisfactory. As a result, greater promotion of this strategic document emerges as a forthcoming task.

Based on these surveys, it can be concluded that the needs are better covered by implemeted infrastructure systems (there were no interviewees who feel that not even the most basic needs are covered), while the coverage of the basic needs by introduced applications is not satisfactory (6% consider the applications do not cover even the basic needs).

The systems *eArchive, eEUCompetitions* and *ePublicProcurements* were identified as crucial for the activities of provincial authorities, even though they were not introduced and the relevant legal framework is not yet completed.

IV. CONCLUSIONS

The document titled "Strategy of eAdministration of Provincial Authorities with the Action Plan until 2015" was drawn up based on:

- initially set goals in 2007,
- data on status of implementation of eAdministration initiatives in the period from 2007 to 2013,
- re-organization of the provincial administration (carried out in the period from 2010 to 2012) and
- stated demands of provincial authorities.

Specifically for the purposes of preparation of the new strategic document "Strategy of eAdministration of Provincial Authorities with the Action Plan until 2015", surveys were conducted to examine:

- level of awareness of top management and employees about the existence and content of the initial document "eAdministration Strategy of Provincial Authorities" (adopted in 2007),
- estimation on the coverage of needs of the provincial authorities by the basic applications and infrastructure systems implemented according to initial document "eAdministration Strategy of Provincial Authorities" (2007),
- assessment of significance of all systems which were not implemented according to plan adopted in 2007,
- current needs of the provincial authorities for ICT support.

Proposals from different provincial authorities regarding new projects were also collected.

The inevitable conclusion is that PDCA (Plan–Do–Check–Act), an iterative four-step management method for the control and continuous improvement and adjustment of processes, can be successfully applied in the process of implementation of eAdministration initiatives and adjustment of action plans [12], [13]. Data collected from conducted surveys were subjected to analysis, which gave important guidelines for preparation of the new Action Plan (until 2015).

This new strategic document will enable the Provincial Government to improve and facilitate development of information society in the AP Vojvodina and introduction of eAdministration at provincial level. The adoption of "Strategy of eAdministration of Provincial Authorities with the Action Plan until 2015" has confirmed Government's commitment to specifying the use of ICT as a priority and as a basis for modernization and increasing the quality of work of Provincial Authorities.

REFERENCES

[1] V. Kumar, B. Mukerji, I. Butt and A. Persaud, "Factors for Successful e-Government Adoption: a Conceptual Framework", *The Electronic Journal of e-Government*, 5(1), 63–76, 2007.

[2] European Commission, "eGovernment Benchmark Framework 2012–2015", Method paper, July 2012. Available at: http://ec.europa.eu/digital-agenda/sites/digital-agenda/files/eGovernment%20Benchmarking%20method%20paper%20published%20version_0.pdf

[3] G. Rudić, B. Dimić-Surla and M. Paroški, "Public library service for accessing records from the AP Vojvodina government sessions", *Journal of Mathematics*, vol. 43, no. 1, pp. 23-31, Novi Sad 2013.

[4] B. Dimić, B. Milosavljević and D. Surla, "XML schema for UNIMARC and MARC 21", *The Electronic Library*, vol. 28, no. 2, pp. 245-262, 2010.

[5] D. Boberic and D. Surla, "XML editor for search and retrieval of bibliographic records in the Z39.50 standard", *Electronic Library*, 27(3), pp. 474-495, 2009.

[6] B. Dimic and D. Surla, "XML editor for UNIMARC and MARC 21 cataloguing", *Electronic Library*, 27(3), pp. 509-528, 2009.

[7] B. Dimic, B. Milosavljevic and D. Surla, "XML schema for UNIMARC and MARC 21", *Electronic Library*, 28(2), pp. 245-262, 2010.

[8] B. Milosavljevic, D. Boberic, D. Surla, "Retrieval of bibliographic records using Apache Lucene", *Electronic Library*, 28(4): 525-539, 2010.

[9] B. Milosavljevic and D. Tešendic, "Software architecture of distributed client/server library circulation system", *Electronic Library*, 28(2), pp. 286-299, 2010.

[10] M. Paroški, V. Popovic and Z. Konjović, "eCompetitions - secure common multilingual electronic public service for support of open competitions for funds in the Autonomous Province of Vojvodina", Proceedings of the 4th International Conference on Information Society and Technology - ICIST 2014, Kopaonik, 9-13 March 2014, pp. 114 – 119, ISBN 978-86-85525-14-8.

[11] M. Paroški, Z. Konjovic and D. Surla, "Implementation of e-Government at the local level in underdeveloped countries: The case study of AP Vojvodina", *Electronic Library*, 31(1), pp. 99-118, 2013.

[12] R. Moen and C. Norman, "Evolution of the PDCA cycle", 2009. Available at: http://pkpinc.com/files/NA01MoenNormanFullpaper.pdf

[13] A. Candiello and A. Cortesi, "KPI-Supported PDCA Model for Innovation Policy Management in Local Government", *Electronic Government*, Lecture Notes in Computer Science, Volume 6846, pp 320-331, 2011.

A strategic approach to providing cloud services for research and education community

Slavko Gajin*, Robert Hackett**, Fulvio Galeazzi***, João Pagaime****
* University of Belgrade, Belgrade, Serbia
** HEAnet - Ireland's National Education and Research Network, Dublin, Ireland
*** GARR - Italian National Research and Education Network, Rome, Italy
**** FCT - Foundation for Science and Technology, Lisbon, Portugal
slavko.gajin@rcub.bg.ac.rs, robert.hackett@heanet.ie, fulvio.galeazzi@garr.it, jpsp@fccn.pt

***Abstract*— The emergence of cloud services and technologies is providing the education and research community with many new and innovative ways to help institutions meet their user's requirements. Cloud services bring many opportunities and benefits but are complex and present many challenges and risks. A well-defined strategy is essential to define the right approach to navigate through the many options and decisions required to provide successful cloud services. In this paper we do not propose a strategy for adopting cloud services for research and education institutions. Instead, we propose a methodology to guideline policy makers in research and education institutions to define an effective strategy to help ensure success in adopting cloud services.**

I. INTRODUCTION

The explosion of the cloud phenomenon over recent years, has provided users with the possibility to acquire and consume high-quality services in new and innovative ways, while at the same time bringing the need to handle an ever increasing volume of data produced and exchanged using an increasing number of devices.

Many research and educational institutions are already using public cloud services (such as email, collaboration and productivity tools) extensively, some have implemented local cloud provisioning, some are planning to deal with cloud services in the near future, while others are still unaware, uninterested or puzzled about the potential of cloud computing.

For the above reasons, it is important for each institution to define its approach to dealing with cloud computing and position its role in providing users with cloud services for research and education in the form of a medium-term strategy. This is a process which requires thorough analysis, planning and decision making, most probably resulting in changes in the organization (skills, technology, business model, working practice etc.) and possibly significant investment in time and money with long term consequences.

The motivation of our work was to help research and education institutions to successfully adopt and provide cloud services for they users. We achieve this by proposing a comprehensive methodology conducted in three stages: firstly the analysis of relevant information, secondly a synthesis phase to identify strategic goals and make crucial decisions, and finally addressing the implementation issues.

The rest of the paper is organized as follows. Section 2 briefly highlights the related work in the field of cloud computing. Section 3 describes cloud strategy analysis process, while Section 4 addresses synthesis of cloud strategy. Section 5 deals with the implementation issues while Section 6 gives the main conclusions of the study.

II. RELATED WORK

In many policy documents cloud computing is recognized as an emerging paradigm that can facilitate innovative research and education. The Digital Agenda for Europe [1] underlines the need to develop EU-wide strategy in order to adopt cloud computing, while the e-Infrastructure Reflection Group (e-IRG) in its study [2] makes the recommendation to establish and promote the necessary policy, rules and legal framework at national and European level. The review of the organizational and institutional implications of cloud computing in higher and further education [3] analyzes the changes to institutional governance, policies, procedures and skills required by adoption of cloud computing. The cloud computing toolkit [4] gives a practical guidance for outsourcing information to the cloud as a starting point for the development of a cloud strategy for higher education institutions. Using cloud computing in higher education is also analysed in [5] as a strategy to improve agility in the current environment caused by the global financial crisis.

The pan-European research and education network – GÉANT [6] that interconnects Europe's National Research and Education Networks (NRENs) and connects over 50 million users at 10,000 institutions across Europe, has aligned its effort with the EC cloud vision. It is achieved by establishing GÉANT's Service Activity group - SA7 Support to Clouds. The recent survey carried out by the SA7 Cloud Strategy Task group has clearly identified that the majority of NRENs perceive cloud services as a fundamental shift which NRENs must recognise and adapt their business, service and even organisation models to accordingly.

This paper presents the results of GÉANT's SA7 Cloud Strategy Task group delivered through GN3plus FP7 project. It is policy oriented paper, with its purpose being to help decision makers in the research and education community to successfully provide cloud services for their users. This strategic approach to cloud computing can be applied at a national level to cover the broad research and education community, targeting funding bodies (research councils, ministries) and e-Infrastructure organizations,

such as NRENs, GRID/HPC initiatives, and nationally operated computer centres.

III. CLOUD STRATEGY ANALYSIS

In the cloud analysis phase, a wide range of information must be gathered and analysed in order to define strategic goals. We propose an analytical process which involves the following steps:

- Understanding the values of key stakeholders (research and educational institutions, individual users, ministries, funding bodies, the NREN);
- Analyse user business processes, needs and demands in relation to the potential of cloud services;
- Analyse the drivers and benefits of using various cloud services from the user perspective;
- Analyse barriers, potential risks and other issues which need to be resolved in order to exploit the benefits;
- Analyse external influences, consisting of many aspects, such as political, economic, social, technological, legal and environmental;
- Analyse internal influences, addressing the institution capabilities to implement the strategy, including internal strengths and weaknesses.

For successful adoption of cloud services the analysis of the above issues from the user perspective is of the most importance, which is described in the rest of this section.

A. User needs and demands

The user community in a typical institution will generate a demand for general computing, network, storage, and application resources to meet common requirements such as email, office productivity applications, video-conferencing, storage, file print and share, CRM, database hosting, web-hosting and the typical needs common to most organisations.

In the educational and research community there is also a significant additional need for more specialized computing and storage resources, that is driven by use-cases specific to the nature of work carried out by the sector. The differences found between the requirements of the two user communities are shown in Table I.

TABLE I.
DIFFERENCES BETWEEN GENERAL AND SCIENCE COMPUTING

General computing	Science computing
Requirements are common to many organizations	Requirements, like performance, are specific
Load varies on daily and weekly cycles e.g. low night use.	Load is high during scientific experiments that can take weeks
Availability may be critical for business normal functions	Availability isn't usually critical (experiments can be restarted a day later in the near future)
Long term predictable use-cases (stable configuration and requirements)	Configurations may vary dramatically according to running experiments

These two user environments have very different requirements and possibly different solutions, so in defining a strategy for cloud services, it is recommended to carefully consider these different requirements.

B. Advantages and benefits

The list of potential advantages and benefits of using cloud computing is outlined below. It is a list of the generally accepted reasons for adopting cloud computing services, but needs to be analysed in relation to the specific business processes in research and education context.

- Cost effective – One of the most attractive benefits of cloud computing is the potential for significantly reducing capital investment requirements. From the perspective of research projects the financing is well balanced, moving the funding from capital investment to operational cost (from Capex to Opex) and exploiting "pay-as-you-go" and "on-demand" payment models.
- Easy and fast deployment (more agility) - The researchers can focus more efficiently on pure research and scientific activities and innovation. In the educational area, for example, students can also easily and quickly obtain the resources to complete their educational activities and release it after that.
- More flexibility and scalability – Research projects and scientific experiments often require capacity which is not always predictable. In a cloud based environment this is not a concern as users can easily increase and decrease capacity.
- Ease of use and access – Simplified usage and universal access from any location with Internet connectivity resulting in improved productivity in the science and research areas, and more efficient learning environments.
- Improved research collaboration – With improved accessibility and data sharing in real time from any location, collaboration is improved both internally in the institution and externally with other local or international partners, such as for example a pan-European project consortium (for instance in the context of Horizon 2020).
- Energy efficient – The adoption of cloud computing results in more optimised usage of computing resources which leads to reduced power consumption, contributing to the greening of the global ICT world.
- Business continuity – Instead of institutions investing in their own disaster recovery facility, significant cost saving can be achieved using external cloud services, which inherently provide high availability and reliability.
- Internal IT transformation – The research and education institution can lower the operational cost for IT maintenance or rather shift the IT focus from system/service administration and maintenance "keeping the lights on" to more valuable tasks such as innovation and support to core research and education processes.

C. Barriers and challenges

Cloud computing as a new technology also brings new challenges, barriers and risks that need to be identified and considered in the cloud strategy and, if possible, be resolved in the implementation phase in order to exploit the benefits of cloud services.

- Security – Security is the biggest concern of most organisations in the adoption of cloud services. The Cloud Security Alliance (CSA) has identified 14 domains to be addressed as part of its security guidance in Cloud Computing [7]. From the users' point of view the areas of biggest concern are the following:
  - Legal and compliance challenges, such as security breach disclosure laws, regulatory requirements, privacy requirements, international laws, intellectual property etc.
  - Information management, including data confidentiality, integrity, and availability, data protection, especially protection of personal data.
  - Identity and access management – since the cloud services are accessible from anywhere and usually are organized in a multitenant environment, users are concerned about how identity and access protection is provided and managed by cloud providers.
- Compliance with existing policies – Using cloud services often involves the outsourcing of sensitive information to the provider's physical location. The concern is significantly higher if the cloud infrastructure is located in a different country under a different legal jurisdiction.
- Lack of control – Moving information and processes to the cloud may involve a significant part of existing responsibilities and control being transferred to the cloud provider. In general, the higher the cloud service is in the deployment stack (going up from IaaS to SaaS) the less control remains with the user over the information management.
- Resistance to new working practices – Most of the technological changes have a positive impact however some resistance to new ways of working may occur, as it may require new roles, responsibilities and skills.
- Skillsets and resources - Depending on the cloud solution e.g. to build or buy cloud services, new skillsets and resources may be required, as the following: legal and contractual expertise, service management, technical expertise, security management, billing and commercial.
- Internal IT transformation – Aside from lowering the cost of IT operation by adopting cloud services, losing in-house IT skills, experience and capacity built up over time, in the long run could lead the institution into a position that would be very difficult and expensive to revert back.
- Integration with in-house systems – Cloud Services in general need to co-exist and integrate with other established IT systems (networks, management & monitoring, backup systems, security systems, federation etc.), which in many cases can be a complex task.
- Technology immaturity - Cloud technologies are evolving rapidly, but are generally regarded as immature, except in the case of some of the major SaaS providers.
- Vendor lock-in - Cloud computing is still faced with the lack of standardization and readiness of commercial cloud providers to fully support interoperability. The users are exposed to the risk of being locked-in to a specific cloud provider with limited or no choices or freedom to move to another one.
- Funding risk - Cloud services in general require complex infrastructure, significant resources and skillsets and consequently significant funding. The startup costs to build a new service may be high and the funding for this may pose a major barrier to progress.

IV. Cloud Strategy Synthesis

Based on the comprehensive set of information, collected and analysed during the strategic analysis process, the next step in cloud strategy formation is to setup strategic goals, which reflect the vision with regard to user requirements. Different business cases, solutions and implementation scenarios need to be investigated to validate if the goals are realistic, feasible and achievable, and therefore if investment is justified. It is an iterative process of strategic thinking with feedback loops where some solutions and options can be, and most probably will be discarded, while new ideas and possibilities will appear. At the end, one or just a few of the more preferred solutions should be identified which represent the best opportunity for cloud deployment with maximum advantages in the most cost effective way.

Setting the goals and making strategic decisions needs to be aligned with the institution's vision and existing strategy and with other policy documents, such as constitutional acts, bylaws, management and operational principles. This policy environment differs for many organisations, but there are a number of initial questions in the context of cloud computing that should be raised in order to drive the strategic thinking:

- What does the institution expect of itself?
- What do others (users, funding bodies, wider community) expect of the institution?
- What is the institution hoping to accomplish?
- What is required to move forward and achieve the goals?

There are many answers to the above questions and therefore many possibilities to further develop the cloud strategy. The strategy development needs to be based on the inputs from the previously performed analysis and driven by user requirements and business cases rather than technical challenges. The specific results of this process are the choices and decisions made on the basis of the following key questions:

- Which user community is targeted - ordinary researchers, "long tail of sciences" researchers, teachers, students etc.?
- Which user needs should be addressed - commodity computing or high performance computing, storage or backup service, collaboration and productivity tools, eLearning, file sharing etc.?

- Which cloud service model(s) to choose – SaaS, PaaS, IaaS or some other?

There are a wide range of combinations of the above possibilities and all are focused on which cloud service to provide for the user community.

Once the cloud service is selected, the next focus of the strategic thinking process is how to provide the cloud service, with two essential questions in defining the cloud solution:

- Which cloud deployment model to implement or support – Public cloud, Private cloud, Community cloud or Hybrid cloud?
- What will be the role of commercial cloud providers in cloud service provisioning?

Answering these questions finally leads the institution to the central point of the cloud strategy:

- What will be the role of the institution in the cloud provisioning?

The rest of the section further discusses these topics that will shape the cloud solution.

A. Cloud Deployment Models

A cloud strategy needs to achieve a balance between the benefits and risks for the institution and make a decision between the different cloud deployment model options [8].

1) Public cloud

Public cloud infrastructure is made available to the general public and the service provision is owned by the institution selling cloud services, e.g. commercial cloud provider. A broad range of public cloud services have been already widely adopted by clients, including the NRENs' community (Gmail, Office365, Salesforce, Amazon Web Services - AWS, HP Cloud, Dropbox etc.). Many of these cloud services are offered for free with limited capacity or functionality, which is often enough for individual usage. At the institutional level the commercial usage of public cloud services is an increasing trend in the research and education community.

2) Private Cloud

An institution may wish to consider providing private cloud infrastructure solely to its internal users (researches, teachers, students). An example of these services is IaaS Compute (VMs) and Storage for scientific experiments, or SaaS services such as a Moodle VLE service for students.

3) Community Cloud

As an extension of internal private cloud, research and education institutions can, possibly by combining their efforts, provide a cloud service to the wider research and education community. NREN and other national e-Infrastructure organisations can play a leading role in providing such a community cloud service(s).

The platform for community cloud is typically quite similar to private cloud e.g. OpenStack, VMware vCloud but providing a community service will add extra requirements and complexity, such as:

- supporting a multitenant environment;
- security requirements to segregate and protect different client environments;
- billing/charging functionality and pricing models
- data protection considerations e.g. data privacy, backups, archiving of client data;
- need to integrate with clients' environment at a network or application level;
- service requirements of clients e.g. SLAs, Service Management Resources; and
- support for AAI

4) Hybrid cloud

The Hybrid cloud model could be an extension of the Private and Community cloud model whereby resources and services available in public cloud(s) can be used in a complementary fashion. An example of this is where compute resources in the private/community cloud can be supplemented by those from a public cloud at particularly busy or peak times.

B. Institution role in cloud provisioning

Cloud services are in general highly complex and bespoke and require significant organisational resources for development and support activities. While some of the building blocks of such services are well known and often use core skillsets of IT staff in research and education community, the additional layers of functionality, such as elasticity, self-service, on demand usage and billing, bring new challenges. While many organisations have the skillsets to develop cloud services and support them in a production environment, for many others this is not feasible. These elements will determine the roles in cloud provisioning, which are analysed below.

1) Public cloud user

Going beyond widely used free but limited usage of public cloud services (mostly SaaS), individual researches, on-going projects or institutional departments can easily purchase public cloud services on the global cloud market. Virtual resources can be elastically scaled up to meet demand and released when no longer needed, while the payment is on a utility basis i.e. "pay-as-you-go" as users consume resources (storage, network traffic, VM size, IP addresses etc.). However, using commercial public services at an organization level must be based on the results of the cloud strategy followed by all the considerations which come with such a decision. It opens up the issues of integration with other institutional services and in-house systems, management of information and processes outsourced to cloud, legal and contractual challenges, other security issues etc.

2) Cloud brokerage

Another level of engagement in the provision of cloud services is to act as a broker for third-party cloud services. It means that a single institution, usually NREN or other e-Infrastructure organisation, negotiate attractive deals from leading suppliers on behalf of the wider research and education community. They do not deliver cloud services directly to their own users as a provider, but instead organize, promote or manage cloud services from the commercial service providers to the user community.

Acting at the national level, NRENs are in position to act as an aggregator and take the lead in the field of cloud brokering and cloud middleware infrastructures, and be able to connect the clouds and provide added value to their members. Reusing existing cloud middleware infrastructures, like AAI, are clear benefits for the user community.

Brokering cloud service offerings should provide clear information to the user, about updated and reliable service descriptions and service levels. Where possible a legal framework for public acquisitions should be made available to NREN users, tailored to the specific community requirements. These objectives lead to an unavoidable vendor management requirement and coordination that should also be availed of to promote inter-operability requirements between different public clouds in order to avoid vendor lock-in traps.

In practice, however, there are some limitations: Leading suppliers already know how to sell to enterprises and don't like additional "middlemen", although competing suppliers may be grateful for help entering the market.

C. *Cloud provider*

There are many possibilities for research and education institutions to deliver their own cloud services, acting as a provider for own users, but commercial cloud providers can still play a significant role. Moreover, the options to deliver a cloud service are determined by the roles and responsibilities of these two main actors – the institution itself and a third party cloud provider, relating to the following key issues in cloud service deployment:

- Ownership – who owns the cloud infrastructure, which includes physical assets, licences, supporting hardware etc.?
- Management – who is responsible for cloud infrastructure governance, operations, monitoring, security provisioning, compliances etc.?
- Location – is the cloud infrastructure located on the organisation's data centre (on-premises) or under the responsibility of the commercial provider (off-premises)?

These three dimensions reflect how the roles and responsibilities are shared between the institution and the third party cloud provider as depicted by the following cube model shown in Figure 1.

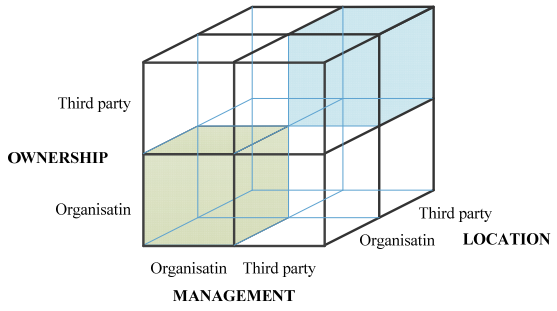


Figure 1: Cloud Provider options

In the extreme case (bottom-left-front corner), the cloud service is fully in-house delivered and provided through a private/community cloud model, developed and managed by the institution itself, running on its own equipment and premises. On the other extreme of the spectrum (top-right-back corner), the cloud service is fully outsourced to a third party provider, who builds and manages the cloud service on its own infrastructure and datacentre.

It is worth to note that in all cases the cloud service is considered to be under the institution ownership and branding.

A safe approach to consider is to use a commercial cloud provider to deliver a dedicated managed service. In this case the institution can use a commercial cloud provider to provide the resources and expertise to build and operate a cloud service, but with the service still being an institution's service. With this approach the cloud provider is effectively providing a managed service. Since a cloud environment is based on multi-layered infrastructure, the roles and responsibilities of cloud infrastructure management in the above cloud provider model can be subdivided and shared across the layers of the cloud stack – from the physical hardware, virtualization platform, up to the application level. The advantages of this approach are that the institution is in a position to provide a new service without having to invest in building up in-house dedicated technical resources and expertise. This could potentially speed up the process of launching new services by avoiding the time, cost and effort of building specialised teams and reduces the financial risk. There are also potential disadvantages with this approach as there is a risk of vendor lock-in as well as increased dependency on a commercial vendor and their ability and commitment to continuing to provide the managed service at a competitive cost.

V. Implementation Issues

A. *Roadmap development*

To ensure that the cloud strategy is successful, a roadmap is needed to define the major activities and resources needed to achieve the strategic goals with acceptable cost, time and efforts. The roadmap would consist of the following three major phases:

- Preparation phase – includes activities to prepare the project team, establish budget and procurement, as well as technical activities to specify technical requirements, acceptance criteria, and design the service with all necessary details to achieve the specification.
- Implementation phase – the activities needed to bring the service live, which includes conducting the procurement, technical installation, configuration, testing and onboarding, as well as supporting activities, such as project management.
- Operational phase – long term activities that includes day to day service operation, monitoring, reporting, maintenance, support, helpdesk, training, promotion marketing etc.

B. *Risk management*

Implementation of a cloud strategy is likely to be challenging due to the potential risks involved regardless to how well the plan is defined and detailed. Cloud strategy therefore needs to include proper risk management in order to anticipate possible risks at an early stage, analyse their impact, and plan mitigation approaches. The goal is to minimize the negative impacts of these unwanted events if they occur, take better decisions and, if possible, turn them into opportunities. To do so, the risk management approach needs to identify possible risks and develop corresponding actions that are incorporated into the initial project plan and budget.

C. Organizational changes

To be able to deliver on its cloud strategy, research and education institution will need to develop the appropriate level of internal competencies to cover the full lifecycle of potential new services from concept to production. This may include some or all of the following capabilities:

- Technical skills in cloud technologies appropriate to the relevant cloud services e.g. platform specific skills to test, implement and operate cloud services, storage, billing etc.
- Security skills to address data protection, identity and access management, compliance to standards and other management and operational security issues.
- Governance, commercial, legal and contractual skills.
- Service management skills to manage an institution's own cloud service or to manage external service providers.

Research and education institutions may have some or all of these competencies in-house but the organisation structure may need to change to reflect the strategy and impact of cloud services on skillsets and resources.

D. Service Branding

The cloud services world is a crowded arena with many service providers competing to get the attention and potential business of institutions. While the research and education community are in general loyal to internal IT infrastructure and services, there is no guarantee that they will choose or even understand the cloud services being offered. The research and education institution will need to compete with strong messages from commercial providers to ensure its users do understand its strategy in relation to cloud services and the key benefits. To ensure its messages reach the right audience it is important to consider:

- The service branding and key messages.
- A communication strategy to deliver the key messages.

The above topics are no different to those facing the commercial providers and the approach is similar. The branding of a cloud service helps users understand its positioning and the unique value that the service brings to research and education community e.g. low cost, high performance, ease of use, security, integration with existing in-house systems (AAI, monitoring) etc.

VI. Conclusion

A strategic approach is essential to ensure successful adoption of cloud services and this paper seeks to help policymakers navigate through the complexity and define a clear strategy for the provision of cloud services. Our study proposes the methodology for the development of a cloud strategy for institutions conducted in three stages.

The initial stage of the strategy development is to conduct an extensive analysis of the needs of both the institution and its users who may have relatively generic and/or highly specific requirements. The analysis stage should also address potential benefits of cloud computing, as well as barriers to the adoption cloud services.

The next stage is the cloud strategy synthesis where the outcome of the analysis stage is used as a foundation to identify strategic goals for those cloud services which are consistent with the organisations' overall strategy and vision. The identification of strategic goals should address the requirements of stakeholders and identify what specific services are required, appropriate service delivery model (SaaS, PaaS, IaaS) and deployment model (Public, Private, Community, Hybrid). We have demonstrated that the role of the institution and the appropriate level of involvement in the delivery of cloud services is a key decision. This role can vary from a full internal develop, build and deploy model to the opposite extreme of a full external managed or brokered service using an external cloud provider and its infrastructure and resources.

The final stage in the development of the cloud strategy needs to focus clearly on how the service(s) should be implemented. It should provide a clear roadmap which defines all the major activities and resources needed to achieve the strategic goals with acceptable cost, time and efforts. The implementation should consider the risk assessment and management as well as the organisational impact such as the requirement for new skills, resources or activities. Finally the institution needs to consider service branding and how it positions and communicates the new cloud service(s) with users.

In summary, cloud services present many opportunities and benefits but also significant challenges and risks. A methodology to define a cloud strategy as outlined in this paper is essential to ensure that right services are identified, designed and delivered using the most appropriate approach that best fits the user and organisation requirements and capabilities consistent with the overall institutional strategy and vision.

Acknowledgment

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7 2007–2013) under Grant Agreement No. 605243 (GN3plus).

References

[1] A Digital Agenda for Europe: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – COM (2010) 245, 19.05.2010.

[2] e-Infrastructure Reflection Group (e-IRG), "Cloud Computing for research and science: a holistic overview, policy, and recommendations", October, 2012.

[3] D. McDonald, A. MacDonald, C. Breslin, "Final report from the JISC Review of the Environmental and Organisational Implications of Cloud Computing in Higher and Further Education", University of Strathclyde, Glasgow, May 2010.

[4] N. Convery, "Cloud Computing Toolkit: Guidance for outsourcing information storage to the cloud", Department of Information Studies, Aberystwyth University, August, 2010.

[5] M. Mircea, A. I. Andreescu. "Using cloud computing in higher education: A strategy to improve agility in the current financial crisis." Communications of the IBIMA 2011 (2011): 1-15.

[6] GÉANT - Pan-European research and education network, www.geant.net

[7] Cloud Security Alliance, "Security Guidance for Critical Areas of Focus in Cloud Computing, v3.0", 2011.

[8] M. Hogan, F. Liu, A. Sokol, J. Tong "NIST Cloud Computing Standards Roadmap, Special Publication 500-291, July 2011

A Contribution to the Development of an Information System in the Function of Improving Decision-Making in Business

Zoran Nešić*, Nebojša Denić **, Miroslav Radojičić*, Jasmina Vesić Vasović*
* University of Kragujevac, Faculty of Technical Sciences, Čačak, Serbia
** Faculty of Information Technology, Belgrade, Serbia
zoran.nesic@ftn.kg.ac.rs, denicnebojsa@gmail.com, miroslav.radojicic@ftn.kg.ac.rs, jasmina.vesic@ftn.kg.ac.rs

***Abstract*— In this paper, the most important elements of software development are presented, with the aim of improving decision making in business. The considered example has practically been implemented in a company engaged in the transportation and storage of petroleum products. The results of the presented software indicate a significant improvement in the collection of large amounts of data and available information. The software support has essentially been focused on the improvement of data visualization for different aspects of the business decision-making process and the business cost reduction. The paper presents the methodological aspects of software development in the key segments. The paper is illustrated with different types of reports with the aim to support managers in business decision making in the specified segment.**

I. INTRODUCTION

The company "Obilić-Petrol", with its seat in Gračanica, is a company for the transport of petroleum products [1]. Within the company, there are five points of sale, i.e. petrol stations, used for the distribution of these goods (Gračanica, Prilužje, Štrpce, Ranilug and Goraždevac). Petroleum products which are sold are recorded through fiscal cash registers. The company has encountered difficulties in collecting and managing information [2]. The company needs software that would help further business decision-making and be of crucial importance for the development and improvement of the quality of the service. In addition, the company would like to be able to import a large amount of data in order to facilitate the processing of integrated data in a correct and uniform manner [3].

Moreover, there is a need for the management of the logistics processes and, consequently, for a possibility of monitoring products' route from suppliers to customers, which is opposed to a bad business organization, a lack of true, accurate and timely information for decision making as well as outdated technologies [4]. All these failures in the company Obilić-Petrol, among other things, have led to a huge deficit in the inventory, adequate control over employees and supplies, a slow flow of information in the enterprise, slow work in the warehouse, which has then also prevented a prompt response to customer complaints, the use of transport routes, the optimal movement of transport means, the optimal storage of huge amounts of paper documents and the like [5]. The current situation in the company makes it impossible for complaints to resolve quickly and the employees to be controlled at the warehouse. The result is reflected in large differences in supplies and a large amount of complaints coming from dissatisfied customers etc.

The information system has resulted in the need for the improvement of the business decision-making process through the reduction of operating costs by creating various types of reports that would assist the management in their making business decisions [6]. Also, data visualization in the company provides greater usability or easier dealing with tasks such as a data analysis, patterns detection and trends prediction. [7] Visualization is helpful in understanding data by summarizing and allowing a faster identification of important parameters, a qualitative review of large and complex databases and a better alignment of the analysis.[8].

The efficiency and effectiveness of small and medium-sized enterprises is improving depending on the quality of processes, and the processes are corrected for improvement efficiency and effectiveness [9]. The visual representation of abstract data is very important in the company when the process of business decision-making is concerned, which increases human understanding (Card, Mackinlay and Shneiderman, 199) [10]. Visualization helps people with large amounts of data and complex problems to obtain relevant information more easily and more quickly [11]. The importance of visualization in the business decision-making process is reflected in the fact that it helps to identify the problem, because it allows the extracting of information from data, a faster insight into the structure of forms and a complete overview of complex information [12]. Decision support systems are of crucial importance for the survival and development of a company, as well as the application of the concepts of information technology in them is [13].

- The development and implementation of a system for multicriteria decision making in business systems [14]
- The improvement of decision-making efficiency by software support [15]
- The improvement of an information system in the function of a business quality [16]
- The application of the modern concepts of management information systems [17]
- The implementation of a broad spectrum of information technology in business [18]

- The implementation of the decision- making system in production planning etc. [19], [20]

The company "Obilić-Petrol" of Gračanica has prepared a thorough analysis of the existing complex processes in the enterprise. Table 1 presents a tabular overview of the existing analysis of costs on a monthly basis, which will form comparative data after the application of software for the improvement of business decision making [21]. Based on the analysis, we can conclude that the company encounters a lack of information for the decision-making process in its operations and that it is one of the key factors of the disorganization and disorder of the storage of oil and oil products. The status quo will result in a negative impact on the relationship between the business partner and the company. Therefore, it was necessary to start the project of the computerization of the enterprises and the storage process as well as the overall logistics process and the renewal of the technological processes that will both ensure the customer's satisfaction with quality services and reduce the cost of the services within its business units [22].

TABLE I.
REVIEW OF THE COST OF THE PREVIOUS SYSTEM [23]

Monthly realization of costs	**EUR**
Labor costs	18,637
Transportation costs	22,700
Material costs	3,197
Costs of inventories	2,300
Other costs	500
Overall costing	**47,334**

The investments necessary for the introduction of software for the improvement of the company "Obilić-Petrol", which include complete IT equipment (hardware, the server, computer, a printer, licenses etc.) amounted to 44,000 EUR for this purpose.

II. DATABASE IMPLEMENTATION

The relational database model is based on the interconnectedness of data stored in tables. In relation to the requirements of the software, the following tables can be defined and will be created:

- The table containing data about the type of fuel, i.e. the petroleum product which is sold at the petrol station
- The table containing data about the workers engaged in the filling of the fuel
- The table containing information on the petrol stations of "Obilić-Petrol Gračanica" and their locations
- The tables containing information on the fuel sold, with supporting data, such as the exact date and time of the sale, the type of the fuel, the petrol station at which it was filled and the name of the worker who filled the fuel.

Based on this, the **Fuel Table** can be defined, containing the following fields:

- **Fuel_ID** – The primary key of the table, which will be auto-incremented and cannot have a NULL value.
- **Fuel_name** - It contains the exact name of the petroleum product.
- **Manufacturer** – It contains the exact name of the manufacturer.

The **Gas_station Table** is also defined and contains the following fields:

- **PStation_ID** – The primary key of the table, which will be auto-incremented and cannot have a NULL value.
- **Station_name** – It contains the exact name of the petrol station.
- **Address** – It contains the exact address of the petrol station.

After that, the **Worker Table** is defined and includes the following fields:

- **Worker_**ID – The primary key of the table, which will be auto-incremented and cannot have a NULL value.
- **Name** – It contains the personal name of the workers.
- **Surname** – It contains the surname of the workers.
- **Birth_date** – It contains the date of birth of the workers.

At the end, the central table – **Petrol_sales Table** – is defined, containing information about the sale of petroleum products:

- **Transaction_ID** – The primary key of the table, which will be auto-incremented and cannot have a NULL value.
- **Fuel_Fuel_ID** – The foreign key of the **Fuel Table**
- **Worker_worker_ID** – The foreign key of the **Worker Table.**
- **Petrol_station_PStation_ID** – The foreign key of the **Petrol Station Table**
- **Sold_litres** – It contains the integral value of the sold litres of petroleum products.
- **Date_time** – It contains information about the exact date and time of the sale.

The relations between the created tables are established, which forms the basis of the relational database model, Fig. 1.

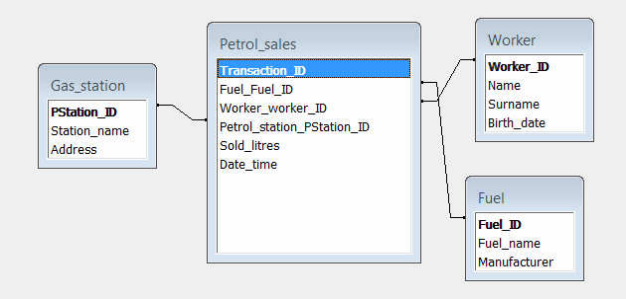


Figure 1. The relational database model [23]

In this way, the tables are successfully created in the database and the relationships between the tables themselves are established. The importance of information management in the organization is assessed by analyzing the ways of using it for the sake of achieving the company's goals. Marchand [24] asserts that, if an organization uses information for management, it can

achieve the following: adding value to products and/or services, improving the operation of risk management, a cost reduction for the business process and for providing goods and/or services to customers and the creation of new opportunities through innovation. If data are organized in different dimensions, i.e. sequences, one can adopt and process larger amounts of information much more easily than when they are not adequately organized and presented [25].

III. GRAPHICAL USER INTERFACE

The primary function of a graphical user interface is the interaction between users and computers by using various graphic elements and text message notifications. When opening the application, the user comes to the main form which gives him the ability to [26]:

- Administer data in the Fuel, Worker and Petrol_Station tables;
- Import the bulk of fuel sales data in the Fuel Sales Table;
- Generate reports required for the management decision support;
- Get basic information about the software; and
- thebe offered a possibility of shutting down the software and exiting the graphic user interface.

The organizational chart of the graphical user interface window and the main interaction of the user interface window itself are shown in Fig. 2.

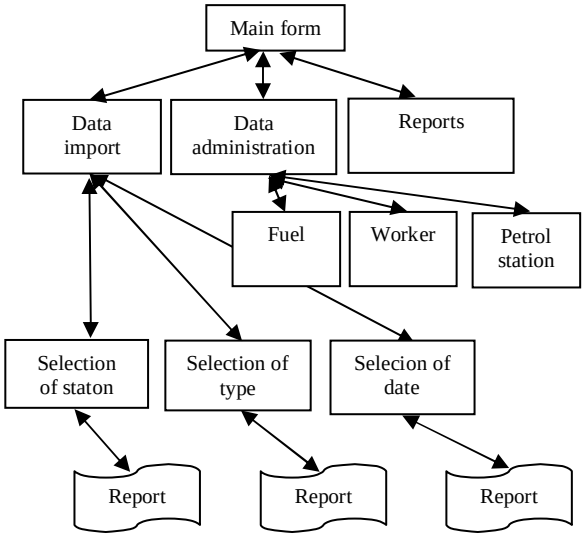


Figure 2. The organizational chart of the user interface [23]

The main form has been designed as an initial screen at the start-up, which has the ability to add and modify data in a database in order to perform a mass import of data into the database and generate reports. [27]

The main form can be seen in Fig. 3.

Figure 3. The main form [23]

The form intended for the administration of data allows the selection of the table in which one wants to add or modify data. On the basis of the above-defined table, it is possible to:

- Administrate information about the Petrol Station,
- Administrate information on the type of the fuel, and
- Administrate data about the workers.

IV. THE MASSIVE DATA IMPORT

As the software will manipulate large amounts of data or, more precisely, manipulate each transaction, where transaction represents the retail sale of petroleum products, it is necessary to adjust the software and the database to the mass import of data [12].

Data will primarily be imported from cash registers, so a flexible way is needed in order have such data transformed from fiscal cash registers into a format easily recognizable to databases and imported into the Fuel Sales Table. For this purpose, the transformation of data from fiscal cash registers into the CSV (Comma-Separated Values) format text will be used so that later it could be possible to import them into a particular database table from the textual CSV format.

The CSV is a batch file where tabular data (numbers and text) in the plain text format are stored. The plain text format represents a string; no data could be interpreted as a binary number. The CSV file consists of a number of production records, where each record stands for a single line; each record consists of fields, and each field is separated from adjacent fields in a certain way – a comma (,) or a tab are most often used to separate the character. Usually, all records have the same field order.

An example of the CSV file that was used to import data into the Fuel Sale Table is displayed in Fig. 4.

```
Sold_Liters, Date_and_time, PStationID, FuelID, WorkerID
98,9/6/2012,2,2,4
126,5/24/2012,2,5,3
122,5/6/2,12,3,2,4
39,9/12/2012,3,5,2
4,4/13/2012,4,4,1
5,10/11/2011,3,6,1
109,12/26/2011,2,2,4
```

Figure 4. An example of the CSV file: [23]

In the form for a massive import data, there is a textual field impossible to edit. This field will get the value of the path to a particular CSV file, where data are imported into the database.

The first step towards an import of data is the selection of the CSV file via the Windows Browse Dialog Box (Figure 9).

By using this option within the VBA code in the onClick event, the next software routine will be applied:

```
Private Sub cmdPath_Click()
Dim FormDialog As FileDialog, vrtSelestedItem As Variant
Dim strSelectedFile As String
Const CSVFAJL = "CSV Files"
Const CSVTIP = "*.csv"
 Set FormDialog =
Application.FileDialog(msoFileDialogFilePicker)
With FormaDijaloga
 .AllowMultiSelect = False
 .InitialView = msoFileDialogViewDetails
 .Filters.Clear
 .Filters.Add CSVFAJL, CSVTIP
  If .Show = -1 Then
   For Each vrtSelestedItem In .SelectedItems
    strSelectedFile = vrtSelestedItem
   Next vrtSelestedItem
   Me![txtPath] = strSelectedFile
  Else
  End If
End With
 Set fdg = Nothing
End Sub
```

With this software routine, the ***ppSelectedFile-of-the String-Type-initialized*** variable is located next to the path of the CSV file, with a limitation that only files with the CSV extension can be selected. After that, out of the ppSelectedFile variable, which at a given moment contains information about the path to the CSV file, those values are assigned by the tkt Path, which in this case represents a text box on a form for a mass data import. After selecting the file, the following VBA programming routine is executed:

```
Private Sub Command4_Click()
Dim strSelectedFile As String
strSelectedFile = Me![txtPath]
DoCmd.TransferText acImportDelim, , "FuelSale",
strSelectedFile, True
MsgBox "Data are succesfully imported."
```

In this case, the value is taken from the text field where the data on the location of the CSV file was previously stored. That data is assigned to the variable ppSelectedFile, after which theDoCmd TransferText procedure is started, transferring the data from the database on the given path, recognized as the CSV file data, into the selected Fuel Sales Table. At the end of this process, the user should be shown a brief piece of information that the correct import of data has token place.

V. REPORTS FOR DECISION-MAKING SUPPORT

Changes that can be caused by high-quality information can easily be limited by the manner of their implementation within a specific decision-making process. [27] When selecting options for reports for decision support in the General Form, the software will open a form where the user is offered to generate six various reports.

Few (2007) argues that a visual display of data is very important because it can, for example, easily differentiate the forms presented in a certain way from other, different patterns displayed. [28]

When selecting options for data processing and report generation, the user will be allowed to enter input parameters for a particular report, after which he/she will receive a generated report.

Eppler and Burkhard (2004) suggest that visual information combines the visual presentation of information and the dynamic technique of receiving and analyzing information by the user. Input data organized in different dimensions make it possible for one to accept and process a larger amount of information than if they were not organized. [25]

A Report on Fuel Sold in a given period of time provides information on the basis of input parameters, such as the start and the end dates, the quantity of the fuel sorted according to the type sold each day between the entered start and end dates.

A Graphic Report on Fuel Sold in a given period of time provides us with information in the form of graphical charts based on input parameters, such as the start and the end dates, the quantity of the fuel that, according to its type, has participated in the total sales of petroleum products between the entered start and end dates.

Valle (2006) states that visualization allows the detection of otherwise imperceptible things – in this argument, we can conclude that the higher the quality of visualization, the easier it is to perceive otherwise unnoticeable parameters and to quickly become aware of the essence, while in the case of low-quality visualization, the user is not provided with an insight into the essence or the existence of information [29].

A Report on Fuel Sold Per Employee on the basis of the entered quantity in a given period of time provides information on how the worker has sold the fuel on the basis of input parameters, such as the start and the end date, as well as the minimum amount of the fuel that will appear in the report. A Graphic Report on Fuel Sold is based on the type of fuel at the petrol station and provides information on the participation of each petrol station in the sale of a certain type of petroleum product on the basis of input parameters, namely the type of the fuel.

A Report on Fuel Sold is based on the definition of the type of the fuel and provides information on how much of a certain type of oil derivative the worker has sold in a certain period of time based on input parameters, such as the start and the end dates and the type of the fuel. A Graphic Report on the Average Amount of Fuel Sold at a petrol station reveals information on the average amount of filled litres per transaction for each type of the fuel sold at the petrol station based on input parameters namely the name of the actual station.

The software used in actual implementation supports the generation of different types of reports:

Report on Fuel Sold in a Given Period of Time: The report gives detailed information on how much of any type of fuel is sold each day within input parameters such as the start and the end dates. A concrete example of this report can be seen in Fig. 5.

Figure 5. A Graphical Report on Fuel Sold in a Given Period of Time [23]

Report on Fuel Sold Per Employee on the Basis of the Entered Quantity in a Given Period of Time: The report is generated for a certain period of time, which time is defined by the start and the end dates and is based on the minimum quantity entered in the form of data in litres, the number of petroleum products sold by each worker, the minimal amount of such data being defined by the user in the form of an input parameter. A concrete example of this report can be seen in Fig. 6.

Figure 6. Report on Fuel Sold per Employee on the Basis of the Entered Quantity in a given Period of Time [23]

Report on Fuel Sold by Employees on the Basis of Defining the Type of Fuel: The report is generated for a certain period of time, which time is defined by the start and the end dates and is based on the type of a petroleum product and the quantity of a particular type of petroleum product sold by each of the registered workers. A concrete example of this report can be seen in Fig. 7.

Figure 7. Report on Fuel Sold by Employees on the Basis of Defining the Type of Fuel [23]

Table 2 accounts for the results of a specific measurable impact on improving the business software in the company "Obilić-Petrol" Gračanica through the improvement of the business decision-making process and a display of the reduction of operating costs.

TABLE II
Display of the costs and measurable results of the software application for business improvement [17]

Monthly realization of costs	EUR
Labor costs	14,800
Transportation costs	23,000
Material costs	3,100
Inventory costs	200
Other costs	500
Total costs	**41,600**

VI. CONCLUSION

In the context of the global economic crisis and the more open world market, companies in Serbia are faced with extremely strong competition. Customer requirements are more demanding, seeking quality services, requiring shorter delivery deadlines, the accurate delivery of goods and the fulfillment of contractual obligations in full. In addition to solving the aforementioned problems, the company "Obilić-Petrol" is making efforts to have its overall business improved by applying software solutions through the improvement of the business decision-making process in order to minimize its labor costs, accomplish the optimal distribution of goods, ensure the adequate quality of its services and increase work efficiency. For this purpose, the company "Obilić-Petrol" has decided to use the concept of the development of information systems for decision support, which, among other things, allows the complete reengineering of the entire logistics process, stressing that in the future, we will not be able to monitor market changes without the use of modern technology and sophisticated tools for managing logistics processes.

After having performed a fundamental analysis, using the concepts of intelligent decision support systems, the reorganization has been conducted in the procurement and distribution of the assortment, changes in transportation have been made and refuelling for particular petrol stations has been done (on the basis of demand and fuel sales); also, a redistribution of the employees has been performed according to the needs of a particular petrol station (workplace).

Having applied the concepts of intelligent decision support systems in such an exact manner, upon receiving the financial report, and after the implementation of the IS, the company for fuel trading "Obilić-Petrol Gračanica" has increased its sales of fuel and additional items, reduced the number of the working hours of employees, i.e. increased the company's profits and improved the quality of the services provided.

Some of the key advantages of this model are:

- Reduction of the cost of storage and distribution.
- Reduction of the time of delivery.
- Increased accuracy of receipt and dispatch of goods.
- Increased worker productivity.
- Increased utilization of office space.

- Improvement of inventory control.
- Management of modern standards of supply chains.
- Customer satisfaction.

The cost and value analysis of the total investment shows that the total return of investment is mostly affected by the reduction in the share of labor costs (the reduction of the employees from 15 to 12) and by a significant reduction in the cost of inventory shortages. The total investment in information technology indicates that the investment itself is fully justified. That is clearly evident through the tabular and graphical reports showing revenue growth, earnings growth, lower interest rates and amortization and net earnings growth in the enterprise "Obilić-Petrol Gračanica" compared to the previous year.

REFERENCES

[1] N. Denić, "Project management and key success factors", May conference on strategic management" IMKSM, Technical Faculty, Bor, pg 480, 2011.

[2] N. Denic, and N. Zivic "Analysis of the factors of ERP solutions implementation in enterprise", Annals of the Oradea University, Fascicle of Management and Technological Engineering, ISSN 1583 - 0691, CNCSIS "Clasa B+", pp 27-31, 2013.

[3] N. Denic, N. Zivic, and B. Siljković, "Project management of information systems", Annals of the Oradea University, Fascicle of Management and Technological Engineering, CNCSIS "Clasa B+", pp 32-36, 2013.

[4] N. Denic, V. Moracanin, M. Milic, and Z. Nešić „Risks the project management of information systems" Tehnički vjesnik, ISSN 1330-3651, Vol. 21 No 6. str 1239-1242 , IF 0,615 za 2013god,

[5] N. Denić, B. Spasić, and M.Milić, " The role of top management in project management implementation of ERP systems", X International May conference on strategic management , Technical Faculty Bor, 2014., pg 3.

[6] N. Denic, S. Markovic, and B. Spasic, "Methodological aspects of ERP (enterprise resource planning) implementation", 14th International Multidisciplinary Scientific GeoConference & EXPO SGEM, Informatics, pg19, June 2014.

[7] J. Zhang, Visualization for information retrieval, Berlin: Springer, 2008.

[8] G. G. Grinstein, and M. O. Ward, "Information Visualization in Data Mining and Knowledge Discovery", V Introduction to Data Visualization San Francisco: Morgan Kaufmann Publishers, pp 21–45, 2002.

[9] D.Chaffey, and S. Wood, ***Business Information Management: Improving Performance Using Information Systems.*** Harlow, England: FT Press, 2004.10 S. K. Card, J. Mackinlay, and B. Shneiderman, "Readings in Information Visualization: Using Vision to Think" Interactive Technologies Paperback February 1999.

[10] C. Speier, and M. G. Morris, " The influence of query interface design on decision-making performance", MIS Quarterly, Vol. 27, No. 3, pp 397–423, 2003.

[11] R. B. Dull, and D. P. Tegarden, Visual representations of accounting information. In: Anandarajan, M., Anandarajan, A., and C. A. Srinivasan, " Business Intelligence Techniques: A Perspective from Accounting and Finance" Berlin: Springer. pp 149–166, 2004.

[12] M. Radojičić, J. Vesić Vasović, and Z. Nešić, "Application of optimization methods in the function of improving performance of organizational systems", Monograph, Faculty of Technical Sciences Čačak, University of Kragujevac, Čačak, 2013.

[13] J. Vesić Vasović, M. Radojičić, and Z .Nešić, "Development of decision making criteria system for production program in industrial companies", 5th International symposium of industrial engineering SIE, Belgrade, pp. 219-222, June 2012.

[14] Z. Nešić, M. Radojičić, J. Vesić Vasović, "Improvement of decision making efficiency by software support", 6th International Quality Conference, Center for Quality, Faculty of Mechanical Engineering, University of Kragujevac, pp. 537-542, June 2012.

[15] L. Ljubić, Z. Nešić, and M. Radojičić, "Improvement of an information system in function of business quality", 6th International Quality Conference, Center for Quality, Faculty of Mechanical Engineering, University of Kragujevac, pp 543-550, June 2012.

[16] M. Radojičić, Z. Nešić, and J. Vesić Vasović, "Some considerations about modern concepts of management information systems", XXXVII Simpozijum o operacionim istraživanjima, SYM-OP-IS, Tara, pp 291-294, September 2010.

[17] M. Radojičić, J. Vesić Vasović, and Z. Nešić, "The development of software support for production management", Monograph, Technical Faculty, Čačak, 2010.

[18] M. Radojičić, Z. Nešić, and D. Ranđić, "Some possibilities of multicriteria optimization in production planning", Technics, special edition, Union of Engineers and Technicians of Serbia, Belgrade, Menadžment, pp. 101-105, 2012.

[19] B. Bilić, M. Radojičić, I. Veža, and Z. Nešić, "Some considerations on the development of the information subsystem for production planning", Journal of Engineering Management And Competitiveness JEMC, Universiity of Novi Sad, Technical faculty "Mihajlo Pupin", Vol. 1., No 1 / 2., pp. 10-15, Zrenjanin, 2011.

[20] N. Denic, M. Milic, and B. Spasic, "Project management impact during ERP system implementation", XIV International Symposium SYMORG 2014, FON Beograd, Zlatibor, Serbia, pp 965-974, June 2014.

[21] N. Denic, B. Spasic, and M. Milic, "ERP system implementation aspects in Serbia", XIV International Symposium SYMORG 2014, FON Beograd, Zlatibor, Serbia, pp 117-123, June 2014.

[22] The internal corporate documents "Obilić.Petrol" Gračanica

[23] D. Marchand et al., ***Mastering Information Management*** Harlow, Financial Times Prentice-Hall, UK, pp. 295-300, 2000.

[24] M. J. Eppler, and R. A. Burkhard, "Knowledge Visualization. Toward a New Discipline and its Fields of Application". ICA Working Paper #2/2004. University of Lugano, July 2004.

[25] N. Denić, B. Spasić, and M. Milić, "Meticulously research project management ERP system implementation in Serbia", The 2nd International Virtual Conference on Advanced Scientific Results (SCIECONF-2014), Zilina, Slovakia, Vol. 1, No. 1., pp 20-26, June 2014.

[26] H. J. Watson, D. L. Goodhue, and B. H. Wixom, " The benefits of data warehousing: why some organizations realize exceptional payoffs", Information & Management, Vol. 39, No. 6, pp 491–502, 2002.

[27] N. Denić, "Strateško upravljanje rizicima i investicijama u ERP sistem", FON, International Symposium SYMORG, Belgrade, Serbia, pp 2557-2564, May 2010.

[28] S. Few, Visualizing Change, " Visual Business Intelligence Newsletter", September 2007. Available at: http://www.perceptualedge.com/ articles/visual_ business_intelligence/visualizing_change.pdf

[29] M. Valle, " Visualization and art" November 2006. Available at: http://www. isedj. Org /isecon/2007 /2523/ISECON.2007.Segall.pdf http://www.cscs. ch/~mvalle/visualization/VizArt.html

ERP AND COMPETITIVE INTELLIGENCE SYSTEMS IN AGILITY OF ORGANIZATION: A SYSTEMATIC LITERATURE REVIEW

Ružica Debeljački*, Pere Tumbas*, Laslo Šereš*
*University of Novi Sad/Faculty of Economics/Department of Business Informatics, Subotica, Serbia
{ruzica, ptumbas, laci}@ef.uns.ac.rs

***Abstract*—Business failure of the organization may be caused by its inability to adapt its business operations to the changes in the market requirements. This leads to a conclusion that today's turbulent business conditions impose the need to increase the level of agility of the organization. Agility is defined as the ability of organization to adapt to different business conditions. Thus, a solution should be looked for in the IT systems such as Enterprise Resource Planning (ERP) and Competitive Intelligence (CI). The reason why exactly these two systems have been chosen is their focus of analysis. ERP systems are specialized for the analysis of organization's business operations, and CI systems have been developed in order to analyze the environment. Therefore, the organization can analyze all business operations by using these two systems. The aim of this paper was to determine whether a synergy of these two systems can contribute to higher level of organization's agility. This paper shows systematic analysis of literature which aim was to determine the contribution of the ERP and CI systems to the agility of the organization and to consider the possibilities of their synergy when used parallelly in today's conditions of successful business operations. This was achieved by considering 35 chosen (selected) scientific papers. The results of analysis showed that the ERP systems, according to most of the authors, do not contribute to the agility of the organization, and the CI systems are the ones which are the most significant component of agility. This synergy can exert influence on the modification of existing business policy or establishment of a new one which would be based both on the analysis of previous business operations and the analysis of business environment, and as a result the organization would have higher level of agility. Based on the obtained results, this paper offers theoretical basis for further research, particularly empirical research, which aim would be to determine the degree of synergy that the ERP and CI systems have in some organizations and their correlations with accomplished (achieved) level of agility of those organizations.**

I. INTRODUCTION

In modern business, there is often a great gap between information which are necessary for management of an organization and information which one organization has [16]. Creation and adequate use of information are some of the key factors of rational organization management. Implementation of information systems requires integrity in all aspects of business operations in order to ensure right and efficient information flow within the organization. Software which provides integration between business processes and safe information flow within the organization is Enterprise Resource Planning – ERP. Original purpose of ERP system was to provide flawless integration of information, business processes, technology and staff [25]. Reference [14] claim that the ability of the organization to create and share information is a good source of competitive advantage. This means that the ERP systems have positive influence on the increase of organization's level of competitiveness [23]. However, today's turbulent business conditions force organizations to have certain level of agility besides their ability to create valuable information. Agility is an important factor for the organization's success and it is defined as the ability to recognize various market opportunities and to take advantage of them [8, 4, 32]. Agility is obviously becoming an obligatory requirement which is set before the organizations so a research regarding the following question could be done:

Rq1. Does ERP system promote enough agility for organization in today's terms of business?

Evolution of client-server and e-business technology have considerably facilitated the process of communication and integration of different systems [19, 27]. When combined with these two technologies, the ERP systems represent vital IT resource for the organization [7]. Reference [19] believes that, in modern business, environment exerts strong influence so the organization needs to be able to control this influence. For the same reason, it is necessary to integrate the environment in business operations in order to create a more flexible organization which will achieve better business results. It is well known that the ERP systems contain internal data which cannot be used for identification of trends in one business environment, so the solution should be looked for in other information technologies. The technology which focuses on the monitoring of environment is Competitive Intelligence (CI). Above all, the CI enables organizations to monitor the competitive environment with the aim of identifying opportunities or threats which should be integrated in business operations. Reference [28] claims that the CI is a process of the organization's acquiring of competitive advantage on the market by understanding the environment better. Still, having the information about environment and competitors is little without information about internal operations of an organization. Internal information of the organization can be used for the analysis of a business conducted for a given period of time. These analyses can show the extent to which business goals have been achieved. However, actual

results of the organization's business operations can be seen only when compared to the results of other organizations. According to this, it can be concluded that there is a need for application of both ERP and CI systems. Having in mind the reference [19] assertion that the organizations need a system which would provide them a framework for certain analyses of business operations and environment, it is indicative that further research of the following subject is needed:

Rq2. Does synergy of ERP and CI increase agility of organization?

This paper shows an attempt to emphasize the role of synergy between ERP and CI systems in the process of gaining competitive advantage and increasing agility of organization because those two systems reflect the business success of the organization. Some claim that success of the organization depends on how well it knows its business processes and competitors, and how efficiently it manages them [16]. This paper offers some theoretical standpoints for further research regarding the effects of synergy between the ERP and CI systems on the agility of the organization.

II. METHODOLOGY

In-depth analysis of scientific papers written on the subject of possible integration of the ERP and CI systems was carried out for the purpose of this research work. The papers were searched for in Ebsco, ScienceDirect and Emerald data bases. The following requirements were set during the search of data bases: that scientific papers were written in English, that they were reviewed positively, that they were presented in scientific conferences or published in academic journals in the period from 2000 to 2014. The papers that were analyzed contained in their abstract or keywords the following terms: „Competitive Intelligence" and/or „ERP". Also, strings such as „ERP and Agility", „CI and Agility", „ERP and CI integration", „Synergy of ERP and CI", „Enterprise Application Integration" were used in the search. These topics were found in 135 scientific papers. After that, those papers were selected based on the requirement that the scientific paper should contain information about issues relating primarily to the increase of organization's agility. Thirty-five scientific papers were selected based on that requirement and they were included in the further process of analysis. The process of selection of scientific papers is presented in Figure 1.

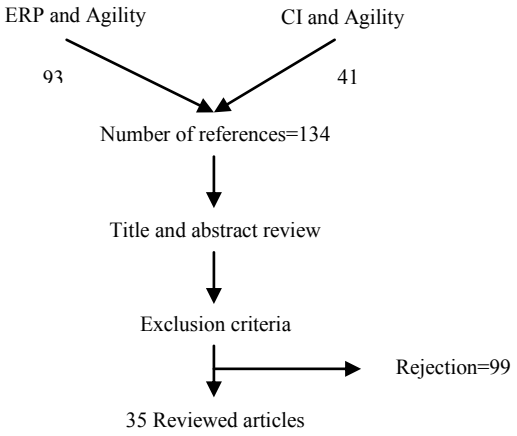


Figure 1. Selection of papers relevant to this research

III. THEORETICAL BACKGROUND

A. Enterprise Resource Planning

Enterprise Resource Planning (ERP) is a software developed with the aim of integrating all business processes of the organization, and by doing this to ensure a comprehensive consideration of business operation using a single data base [19]. It must be noted that the ERP systems are not a uniform solution for all organizations, and they need certain modifications in order to suit different needs [18, 24, 30, 11, 10, 5]. Implementation of ERP systems enables automatization of all business processes of the organization [20, 16, 1, 17] They provide systematization and organization of information about everyday business operations into one data base. This ensures the availability of information which are necessary for further creation of adequate information that in turn support the decision making process [26]. Some authors claim that the ERP systems represent the integration of different applications which aim is to provide support for the realization of business processes and maintain the information flow in the organizations [19, 24, 13]. Moreover, they provide better insight into critical points in the realization of business processes which consequently increases efficiency and productivity of the organization [25, 27]. Based on the given description of the ERP system, it is obvious that this system is focused on internal business process of the organization. However, market uncertainty forces organizations to adapt to different business conditions and changes in the environment. Therefore, organizations need to achieve more flexibility and agility. There is a belief that in spite of high investments in ERP systems, they still do not ensure competitive advantage and agility of the organization [25, 18, 35, 24, 16, 1, 33]. Reason for this may be the fact that the ERP systems were primarily designed to distinguish clearly between the organization and environment in order to minimize its negative influence [14]. On the other hand, reference [16] disagree with that and claim that the true benefit from ERP systems can be gained only by connecting it to the environment which consequently increases agility of the organization. In order to do that, a synergy or integration of EPR system and other systems needs be created [29, 30, 22, 9]. Synergy should enable the organization to use data and information both efficiently and effectively, and to make better decisions which result from adequate response to market changes. Thus, based on the previously said, a synergy between the ERP system and some other technology with strong analytical features for environment should contribute to the success of the organization. Synergy between the ERP and CI systems should provide an integrated functional platform which has direct influence on the decision making process and agility of the organization [6].

B. Competitive Intelligence

Reference [12] think that organizations need to implement the environment in its IT resources in order to survive and advance in today's business conditions. There are also some statements about Business Intelligence (BI) systems becoming increasingly common in the process of achieving more agility [4, 32]. Reference [34] claim that the CI developed into an important field within the BI

which focuses on understanding and measuring of the influence of outside environment on the results of the organization. Some authors believe that only those organizations which have appropriate and coherent information system such as the CI will manage to advance in today's unstable business environment [16, 2]. The CI can be defined as a process of collection and analysis of publicly available information on business environment, competitors and the organization itself, in accordance with the ethics policy [3]. These systems make forecasting of future trends possible, thus increasing the ability of organization to adapt to the changeable conditions of its environment [28]. Thus, efficient application of the CI systems in business operations is the key factor for the organization to develop its agility [28, 15]. Fast decision making and agility depend solely on how progressive is the organization in realizing new opportunities as well as the threats present on the market [16]. Reference [31] say that the application of the CI systems offers exactly that to the organizations. The CI system explores the information related to the environment and competitors from different sources such as the internet, blogs, social and business networks, on-line media and alike. The obtained data are then implemented in current business processes and different modifications of business operations are made. Still, there are some limitations. Considering the fact that the CI systems are focused on the environment and competitors, they primarily use the external data. In order for the organization to survive in an unstable environment and to acquire agility, it is necessary to manipulate with both internal and external data in the decision making processes. This brings us to the conclusion that maximum efficiency of the CI system can be obtained in the cooperation with other information systems.

IV. RESEARCH RESULTS

Market uncertainty imposes constant need for the organization to adapt to the business environment. Reference [16] think that the organizations should be capable of adapting to the changes imposed by business environment in order to survive on the market. The key to success of the organization is its ability to see and react to the mentioned changes [25]. Therefore, agility is the crucial factor for business operations of every organization [21, 22]. Considering the described significance of organization's agility in business operations, Table 1 shows authors of the analyzed scientific papers and their opinions about the influence of the ERP and CI systems on agility.

Authors	ERP	ERP does not provide agility	CI	CI provide agility	synergy of ERP and CI provides agility
	22	6	15	7	8
	62.86%	27.27%	42.86%	46.67%	22.86%
Banker et al (2006)	x	x			
Calof and Wright (2008)			x	x	
Cavalcanti (2005)			x		
Chen and Siau (2012)	x		x	x	x
Davenport, Harris and Cantrell (2004)	x				
Gaidelys (2010)	x		x		x
Gleghorn (2005)	x		x		x
Gong and Janssen (2012)					
Goodhue et al (2009)	x		x		x
Hitt, Wu and Zhou (2002)					
Ketokivi (2006)	x				
Koh and Maguire (2009)					
LaFata and Hofmann (2004)	x				
Lengnick-Hall, Lengnick-Hall and Abdinnour-Helm (2004)	x				
Liu and Wang (2008)			x	x	
Maguire, Habibu and Ojiako (2010)	x		x	x	
Markus (2000)					
Nazir and Pinsonneault (2012)	x	x			
Özkarabacak, Çevik and Gökşen (2014)	x				
Pei-Fang (2013)					
Raschke and David (2005)					
Sambamurthy, Bharadwaj and Grover (2003)	x		x		x
Seddon (2005)	x				
Seddon, Calvert and Yang (2010)		x			
Seethamraju and Krishna (2013)	x	x			
Sharif, Irani and Love (2005)	x				
Swaminathan and Tayur (2003)	x				
Štefánikováa and Masárová (2014)			x	x	
Tallon and Pinsonneault (2011)	x		x		x
van Oosterhout, Waarts and van Hillegersberg (2006)	x		x		x
Vitt, Luckevich and Misner (2002)			x	x	
Xiaofeng and Keng (2011)	x		x	x	x
Wade and Hulland (2004)	x	x			
Zheng, Fader and Padmanabhan (2012)			x		
Zhu, Wang and Chen (2010)	x	x			

Table1. Authors and the subjects examined in their research

Table 1 shows that 6 out of 22 (27.27%) analyzed scientific papers, which were written on the subject of ERP systems, examine their influence on the agility. All authors agree that ERP systems do not contribute significantly to the development of agility. The reason for this is the fact that these systems are focused on internal business operations of the organization, so their strategic role is detecting the critical points in the realization of business processes. This focus on internal business operations of the organization makes them a necessary but insufficient precondition for acquiring the desired level of agility. Organizations have realized the availability of abundance of information within the ERP systems, but the real challenge is to process these data and collect them from the business environment. Collection and processing of data from the environment have come to the center of attention due to the fact that organizations operate at times of high competition (hyper competition). Such business conditions set a requirement for the organizations to do constant monitoring of the environment and competition with the aim to survive and achieve better position on the market. This is supported by Big Data phenomena which considerably influenced the availability of data from external sources. In that sense, the ERP systems should be accompanied with some information technologies which are good for analytical analysis of the environment. IT technology such as the CI system could be a good choice for the organizations. The CI systems enable infrastructural development which helps detecting changes in business environment. Table 1 shows that 46.76% (7 out of 15) of scientific papers written on the subject of the CI confirm that this technology helps in acquiring more agility. On the other hand, there are also some statements about the CI being insufficient by itself for that purpose. The reason for that is that this system does not contain important data on business operations of the organization, like ERP systems do. This means that the application of ERP systems in combination with the CI system can give best results regarding the achievement of desired level of agility of organization, which was also confirmed by 22.86% authors. These two systems can help modify present business policy or create a new one which would be based on the analysis of former business operations and business environment which would, as a result, create more agility of the organization. Synergy of these two systems provides necessary information needed for successful business in modern conditions. Without any of these two, it would be hard to carry out any analyses required for strategic management of the organization. Having in mind that the nature of these two systems is diametrically opposed, since the ERP systems are operational and the CI systems are analytical, integration of these two systems is very difficult. Besides their nature, there is also an obvious difference in the type of data these two systems contain. The ERP systems contain structural data, while the CI systems have non-structural data. In spite of different technological bases of these systems, the integration of data can be achieved by storing them in the Data Warehouse - DW. This way, these systems can be used as the source for BI system which is applied in order to provide efficient and effective analysis of business operations and better decisions. Integration of data from these systems in the DW can give valuable analyses which further enable timely reaction to changes in the environment. DW is actually the result of a synergy of these two systems, and the results of the organization can be assessed only by using the data from both systems. For example, having DW with stored information from these systems can enable qualitative SWOT analysis. Data obtained from ERP system is related to Strengths and Weaknesses (SW), and CI systems provide data on Opportunities and Threats (OT). Feedback of SWOT analysis gives possibility of creating more qualitative information and consequently making better decisions, which emphasizes the strategic role of synergy of these systems. This is only one of the examples of the effects of synergy of these two systems which confirm the significance of their application.

V. FUTURE RESEARCH DIRECTIONS

Results of systematic analysis of literature in this paper give theoretic basis for further research, especially the empirical research, which aim is to determine the level of synergy between ERP and CI systems in come organizations and their correlation with acquired (achieved) agility level. Also, some research in the future should focus on the analysis of those segments of business operations which could benefit the most from the synergy between ERP and CI systems in organization's business operations.

CONCLUSION

ERP systems have simplified, standardized, integrated and automatized business operations, but they have not contributed to the organization in the sense that the organization acquires more agility and can survive in unstable business conditions. It is believed that agility can hardly be acquired by implementing only ERP systems. Most of the organizations have already implemented the ERP system, but great market uncertainty encourages organizations to invest in the implementation and application of the CI systems. The result of synergy of these two systems gives more qualitative information, better decisions, increased productivity, better position on the market, and alike. Having analyzed scientific literature, it was noticed that empirical research on the subject of CI was still scarce. It may be due to the fact that the CI is a relatively new technology. Considering today's unstable business conditions and the fact that the CI technology has been developed to primarily monitor and analyze business environment and competition of the organization then, the significance of its application is obvious. By using CI systems the organization can gain leading market position. Leadership is gained by monitoring the environment in order to determine current trends and opportunities. This way, the organization can ensure its favorable market position which implies the possibility for the organization to be the one that dictates the rules in business and not the one who follows the others. To this end, modification of business policy also needs internal data about business operations of the organization, besides the external data obtained by analysis and monitoring of environment and competition, and those data are provided by the ERP system. This leads to the conclusion that implementation of the ERP and CI systems enable organizations to have proactive business operations. By doing this research of literature, a

theoretical basis was given which emphasizes the importance of the synergy of these two systems..

REFERENCES

[1] Banker, R., Bardhan, I.R., Chang, H., Lin, S. (2006) "Plant information systems, manufacturing capabilities, and plant performance". *MIS Quarterly* Volume 30, 315–337.

[2] Calof, J.L., Wright, S., (2008), "Competitive Intelligence: A Practitioner, Academic and Inter-Disciplinary Perspective". *European Journal of Marketing*, 42(7/8), pp 717-730.

[3] Cavalcanti, E. P. (2005), "The Relationship between Business Intelligence and Business Success". *Journal of Competitive Intelligence and Management*, 3(1), Spring 2005

[4] Chen, X., Siau, K. (2012). "Effect of Business Intelligence and IT Infrastructure Flexibility on Organizational Agility". *Thirty Third International Conference on Information Systems. Orlando*.

[5] Davenport, T. H., Harris, J. G., Cantrell, S. (2004). "Enterprise systems and ongoing process change". *Business Process Management Journal*, 10(1), 16-26

[6] Gaidelys, V. (2010). "The role of competitive intelligence in the course of business process". *Economics and Management*, Volume 15, 1057-1064.

[7] Gleghorn, R., (2005). "Enterprise Application Integration: A Manager's Perspective". *IEEE Computer Society*, pp. 17-23

[8] Gong, Y., Janssen, M. (2012). "From policy implementation to business process management: Principles for creating flexibility and agility". *Government Information Quarterly*, Volume 29, 61-71.

[9] Goodhue, D., Chen, D., Boudreau, M., Davis, A., Cochran, J. (2009). "Addressing business agility challenges with enterprise systems". *MIS Quarterly Executive*, 8(2), 73-87.

[10] Hitt, L., Wu, D., Zhou, X., (2002). "Investment in enterprise resource planning: business impact and productivity measures". *Journal of Management Information Systems* 19(1), 71–98.

[11] Ketokivi, M. (2006). "Elaborating the contingency theory of organizations: the case of manufacturing flexibility strategies". *Production and Operations Management*, 15(2), 215-228

[12] Koh, S.C.L., and Maguire, S., (2009), "Information and Communication Technologies Management in Turbulent Business Environments". Published Information Science Reference, Hershey, USA.

[13] LaFata, J., Hofmann, S. (2004). "Enterprise Application Integration A Primer in Integration Technologies". *Liquidhub: Fueling Business Transformation*.

[14] Lengnick-Hall, C. A., Lengnick-Hall, M. L., Abdinnour-Helm, S. (2004). "The role of social and intellectual capital in achieving competitive advantage through enterprise resource planning (ERP) systems". *Journal of Engineering and Technology Management*, 21(4), 307-330.

[15] Liu, C-H., Wang, C-C, (2008), "Forecast Competitor Service Strategy with Service Taxonomy and CI Data", *European Journal of Marketing*, 42(7/8), pp 746-765.

[16] Maguire, Stuart; Suluo, Habibu; and Ojiako, Udi, (2010). "Competitor intelligence: the real value from ERP II?". *UK Academy for Information Systems Conference Proceedings 2010. Paper 36.*

[17] Markus, M.L., (2000). "Paradigm shifts – e-business and business/systems integration". *Communications of the Association for Information Systems* 4 (10), 1–44.

[18] Nazir, S., Pinsonneault, A. (2012). "IT and firm agility: an electronic integration perspective". *Journal of the Association for Information Systems*, 13(3), 150-171

[19] Özkarabacak, B., Çevik, E., Gökşen, P. Y. (2014)." A Comparison Analysis between ERP and EAI". *Procedia Economics and Finance*, *9*, 488-500.

[20] Pei-Fang Hsu, (2013). "Commodity or competitive advantage? Analysis of the ERP value paradox". *Electronic Commerce Research and Applications*, 12 (6), 412–424

[21] Raschke, R., David, J. S. (2005). "Business process agility". *Proceedings of the 11th Americas Conference on Information Systems*, Omaha, NE, USA, 11e14 August pp. 355-360

[22] Sambamurthy, V., Bharadwaj, A., Grover, V. (2003). "Shaping agility through digital Options: reconceptualizing the role of information technology in contemporary firms". *MIS Quarterly*, 27(2), 237-263

[23] Seddon, P. B. (2005). "Are ERP systems a source of competitive advantage?", *Strategic Change*, 14(5), 283-293.

[24] Seddon, P., Calvert, C., Yang, S. (2010). "A multi-project model of key factors affecting organizational benefits from enterprise systems". *MIS Quarterly*, 34(2), 305-328

[25] Seethamraju, R., Krishna Sundar, D. (2013). "Influence of ERP systems on business process agility". *IIMB Management Review*, *25*(3), 137-149.

[26] Sharif, A.M., Irani, Z., Love P.E.D., (2005). "Integrating ERP using EAI: A Model for Post-hoc Evaluation*". European Journal of Information Systems*, Vol. 14, No. 3, pp. 162-174

[27] Swaminathan, J., Tayur, S., (2003). "Models for supply chains in e-business". *Management Science* 49 (10), 1387–1406.

[28] Štefánikováa, Lj., Masárová, G. (2014) "The need of complex competitive intelligence". *Procedia - Social and Behavioral Sciences* 110 669 – 677

[29] Tallon, P. P., Pinsonneault, A. (2011). "Competing perspectives on the link between strategic information technology alignment and organisational agility: insights from a mediation model". *MIS Quarterly*, 35(2), 463-484

[30] van Oosterhout, M., Waarts, E., van Hillegersberg, J. (2006). "Change factors requiring agility and implications for IT". *European Journal of Information Systems*, 15, 132-145.

[31] Vitt, E., Luckevich, M. and Misner, S. (2002), "Business Intelligence: Making Better Decisions Faster*", Microsoft Press, Redmond, Washington*

[32] Xiaofeng Chen, Keng Siau, (2011). "Impact of Business Intelligence and IT Infrastructure flexibility on Competitive Performance: An Organizational Agility Perspective". *ICIS 2011 Proceedings,* Paper 23.

[33] Wade, M., Hulland, J., (2004). "The resource-based view and information systems research: review, extension, and suggestions for future research*". MIS Quarterly* 28 (1), 107–142.

[34] Zheng, Z., Fader, P., Padmanabhan, B. (2012). "From business intelligence to competitive intelligence: Inferring competitive measures using augmented site-centric data". *Information Systems Research*, 23(3-part-1), 698-720

[35] Zhu, Y., Li, Y., Wang, W., Chen, J., (2010). "What leads to post-implementation success of ERP? An empirical study of the Chinese retail industry". *International Journal of Information Management* 30 (3), 265–276.

Advantages and Drawbacks of Sloodle application for creating high-quality teaching materials with demanding graphics

Maja Radovic*, Danijela Milosevic*, Andjelija Mitrovic**, Marija Blagojević*
* Faculty of technical science/Information technology, Cacak, Serbia
** Technical College Cacak/Engineering, Cacak, Serbia
maja.radovic@ftn.kg.ac.rs
danijela.milosevic@ftn.kg.ac.rs
andjelija.mitrovic@vstss.com
marija.blagojevic@ftn.kg.ac.rs

Abstract— **Due to is potentials, Sloodle virtual environment has been used to create medical course for training medical staff in the field of orthopedic. As opposed to conventional practice, where 3D models of vertebrae and spine are difficult to obtain and expensive for the preparation, Sloodle medical course allows participants access to high quality 3D models in virtual world at any time. The paper presents main features of course design and implementation. Besides, the internal evaluation is performed aiming to analyze the possibilities offered by presented environment in order to determine whether tutors can use Sloodle to obtain high-quality teaching materials with a reasonable investment of time and knowledge. The evaluation included six categories of questions and assessed the Sloodle tools management, upgrading, collaboration, costs and simplicity of adding 3D models, ease of programming, etc. The advantages and drawbacks are discussed in detail.**

I. Introduction

Over the last few years, there has been growing interest in the medical and public health communities for using virtual learning environments (VLE) for education and training [1]. Unlike traditional learning and practice on real patients, virtual environment allows students to learn by making mistakes. Students can also be on remote locations, and still have the opportunity to attend the training. Along with that, the Serbian national project "Applying biomedical engineering in pre-clinical and clinical practice" strives to create the VLE course for training the medical staff in the field of orthopedics, particularly spine column deformities.

Apart from appropriate educational content, the course needs to satisfy different aspects of the training, such as high quality 3D models of different spine column deformities, availability, cost effectiveness, etc. While Learning Management Systems, such as Moodle [2] perform most of these functions, they still limit students to deal with only specified activities and to have virtually no control over the conditions in which activities occur. Such missing functions of the LMSs have been overcome through the contribution of the tools that exist in 3D virtual environments, such as Second Life (SL) [3], which provides a new range of educational opportunities [4]. Both LMSs and SL have necessary functions for learning not exists in the other. Integration of these two systems is presented in the form of Simulation Linked Object Oriented Learning Environment (Sloodle) [5]. Due to its capabilities, we have chosen Sloodle for our course design and implementation.

Sloodle environment provides a variety of different tools, which make managing educational activities in Second Life easier [4]. Some of the tools are used in the medical course for delivering lectures or collecting feedback and assignments related to Second Life activities.

There is large number of research regarding application of VLE in medical education [6], [7], but just a few of them relates to Sloodle [8], [9]. The same environment is applied for trainings in different fields, such as programming languages [9], or Computing Graphics [10]. We didn't find any published research regarding Sloodle application in the field of orthopedics. The overview of course preparation and its Sloodle implementation is described in detail by authors in [11].

II. Course design and implementation

The "Spine – functions and deformities" medical course is implemented on Moodle platform within Faculty of technical sciences Cacak, University of Kragujevac. The virtual Island College of Scripting, Music, and Science, Horsa [12] is selected to be the starting point in Sloodle VLE. Renting or purchasing a land for the permanent course maintenance is planned in the near future. Sloodle module called Sloodle Controller is used to connect Moodle course with Second Life.

We have designed medical course based on principles of instructional design, and according to consultation with medical expert (Fig. 1).

Due to assumption that medical staffs are not familiar with virtual worlds, first topic of the course offers basic guidelines for existence in Second Life, such as creating avatars (a virtual representation of themselves), acquiring appropriate skills, connecting users' avatar to their Moodle account, etc.

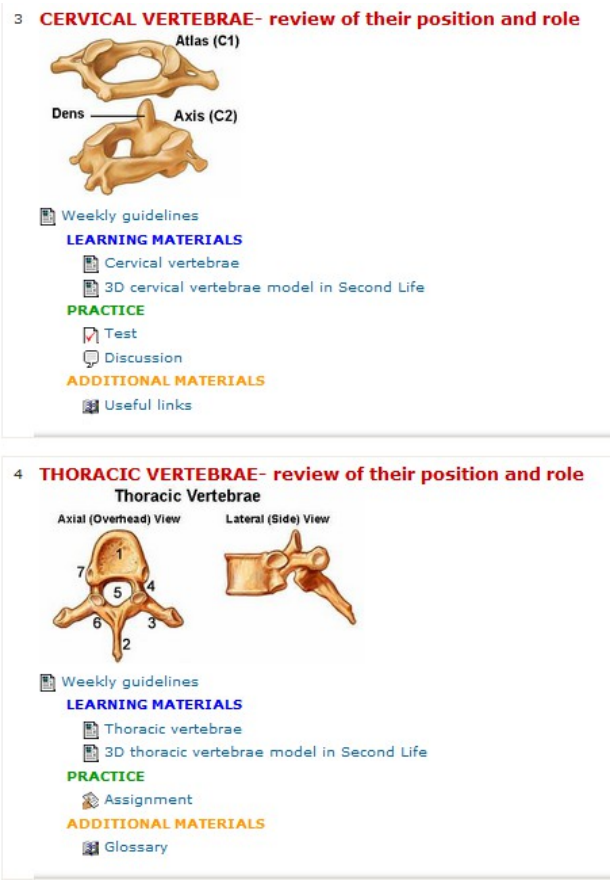


Figure 1. The screenshot of Sloodle medical course "Spine – functions and deformities"

After accommodation to virtual environment, users will recall their knowledge of detailed structure of the spinal column, and through series of activities focuses on different spinal deformities with special emphasis on scoliosis, kyphosis and lordosis.

A detailed list of course topics is the following:

- Second Life – introducing to virtual world
- Spine column – basic concepts
- Cervical vertebrae – review of their position and role
- Thoracic vertebrae – review of their position and role
- Lumbar vertebrae – review of their position and role
- Sacrum and Coccyx – review of their position and role
- Scoliosis – diagnosis, types, treatment
- Kyphosis – diagnosis, types, treatment
- Lordosis – diagnosis, types, treatment
- Comprehensive self evaluation

Topics are organized in three sections: learning materials, practice and additional materials. Learning materials are rich in multimedia resources. Apart from that, learning materials include links to Second Life locations where 3D models of particular vertebrae are placed (Fig. 2).

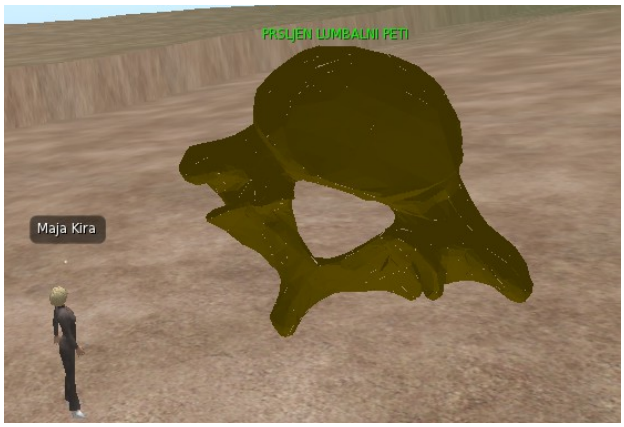


Figure 2. Overview of the fifth lumbar vertebrae in Second Life

Because of its complex structure, reflected in the large number of surfaces, graphic spatial models cannot be directly created or imported to Second Life. They are externally developed in CATIA software, which is the most powerful and widely used CAD (computer aided design) software of its kind in the world. Upon the successful creation of spine parts models, we have experienced the serious problem while trying to upload them in SL. The uploading led to huge number of triangles and vertices in Second Life that cannot be linked again into one model. After the extensive research, we resolved this issue by introducing Blender application. When the model is inserted in program, through series of steps its number of triangles and vertices were reduced (without losing the model quality). The process of uploading edited vertebrae model to Second Life is same as any other object (Fig. 3).

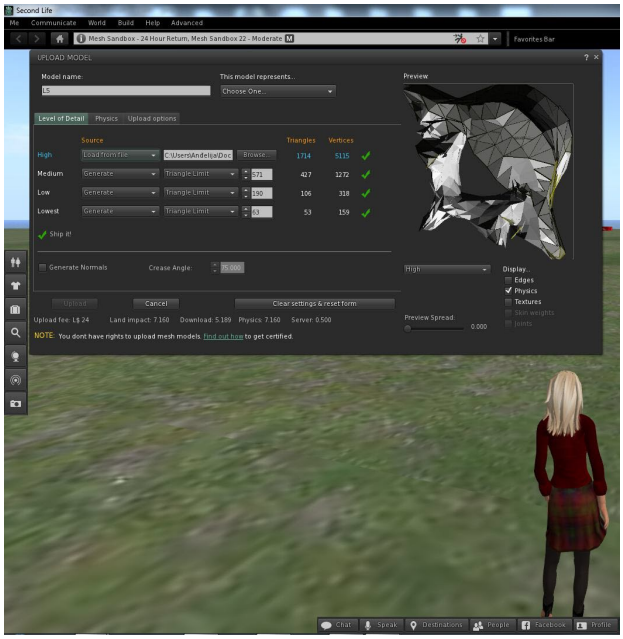


Figure 3. Uploading vertebrae model from Blender to Second Life

In order to additionally enhance models' visualization we have used Linden Scripting Language (LSL), which is official Second Life programming language. The LSL functions are used to add functionalities to spine part models, such as appropriate actions and labels. Fig. 4 shows part of the LSL script used on vertebrae model.

The course practice section contains tests and assignments that can help users to test their knowledge. Additional course materials contain useful links, terms and other assistance to the appropriate topic.

```
default
{
    touch_start(integer x)
    {
        integer side=llDetectedTouchFace(0);
        if(side==0)
    {
        llSetColor(<1,0,0>,0);
        llSetLinkColor(LINK_SET,<1.0,0.0,0.0>,0);
        llSetText("You have touched red surface",<1,0,0>,1.0);
        llTargetOmega(<1,0,0>,PI/4,0.5);

    }

    else if(side==2)
    {
        llSetColor(<0.0,1.0,0.0>,2);
        llSetLinkColor(LINK_SET,<0.0,1.0,0.0>,2);
        llSetText("You have touched yellow surface",<0.0,1.0,0.0>,1.0);
        llTargetOmega(<0,0,1>,PI/4,0.5);

    }
    else if(side==3)
    {
        llSetColor(<0.0,0.0,1.0>,3);
        llSetLinkColor(LINK_SET,<0.0,0.0,1.0>,3);
        llSetText("You have touched blue surface",<0.0,0.0,1.0>,1.0);
        llTargetOmega(<0,0,-1>,PI/4,0.5);
    }
}
}
```

Figure 4. Overview of LSL script

Two categories of Sloodle tools were used during the course design. Enrolment tools manage student access permission for a virtual classroom, help students register at Moodle and enroll them in the appropriate Moodle course.

Educational tools allow students to work with Moodle activities in Second Life. Within each topic we used the following Sloode educational tools:

- WebIntercom. It connects chat sessions in Second Life to Moodle chatroom. Given that this is synchronous activity, we anticipated its usage for chat sessions where user will have discussion on particular issue, such as different treatment of lordosis, etc.

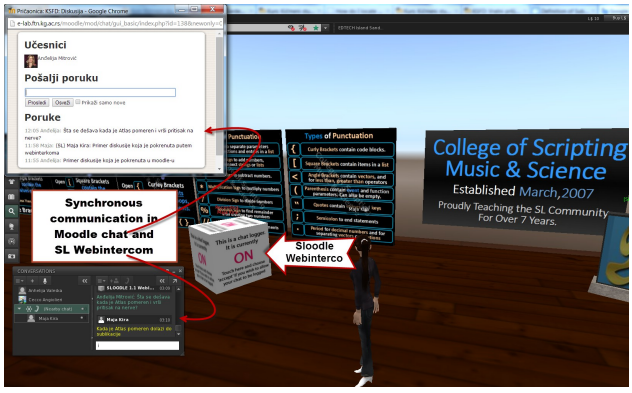


Figure 5. Application of Webintercom tool

- MetaGloss is also one of useful Sloodle tools, which allow access to Moodle glossary. Each topic has glossary, which contains terms used in that topic. For example Thoracic vertebrae – review of their position and role topic contains glossary that contains terms such as Vertebral Body, Spinous Process, Transvers Facet, etc.

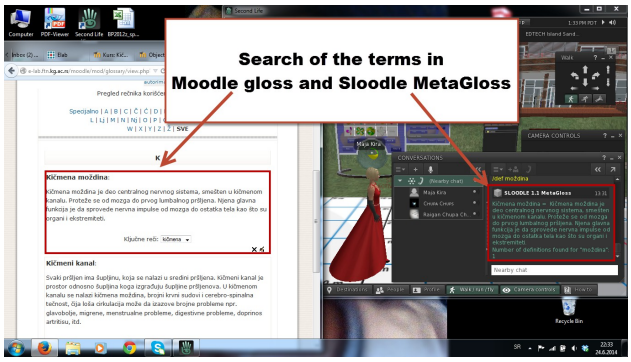


Figure 6. Application of MetaGloss tool

- Quiz is Sloodle tool used for testing. Quiz can be accessed from Moodle as standard test, or from Second Life as Quiz Chair. Primary purpose of quizzes in our medical course is self-evaluation. For example, after processing a topic related to kyphosis, users can do the quiz, which deals with various types of kyphosis and thus test their knowledge. Results from quiz performed in Second Life can be later reviewed in Moodle.

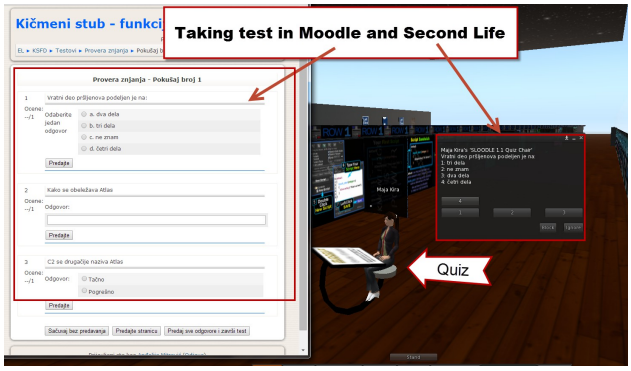


Figure 7. Example of quiz in Moodle and Second Life

- Prim Drop tool allows users to submit Second Life objects in a Moodle database. Tutor can create assignment, which requires users to choose appropriate spine model among several available models. Users will submit chosen model using Prim Drop tool. This Sloodle tool will additionally engage users to hand in assignments in-world.

It should be noted that it is planned to use some more Sloodle tools, which require ownership or renting the land, as envisaged in the near future. Such tool is Sloodle Presenter, which enable participants to view presentations and videos related to the spinal cord, in Second Life, which are previously posted on Moodle.

Before presenting the course to the users, we have conducted internal evaluation with questions about Sloodle tools management, possibilities for upgrading, collaboration, costs and simplicity of adding 3D models, ease of programming, etc. The main aim of evaluation was to determine whether teachers can use Sloodle to obtain useful and high-quality teaching materials with a reasonable investment of time and knowledge.

III. Evaluation of virtual environment course

This section provides internal evaluation of proposed virtual learning environment. The goal of evaluation is the subjective analysis of possibilities offered by presented environment. The response is given from tutors' aspect. The questions are classified in six categories:

- Access to the course and Account
- Structure of course and resources
- Communication tools
- Progress Tracking Tools
- Technical Issues
- Collaboration issues

Table 1 shows the list of evaluation questions and responses that are developed by course designers and technical assistants. The Likert scale is used for responses, ranging from (5 – high satisfaction) to (1 – low satisfaction).

According to the provided response, it can be concluded that, in general, presented environment provide many useful opportunities and enhancements for teaching process.

IV. Advantages and potential drawbacks from teacher's perspective

There are several advantages for choosing Sloodle application to create medical course. Participants can enhance their skills and knowledge through work with different 3D models of vertebrae and spinal cord, which are often not available in real life practice. Also, traditional practice requires participants to be at the same time in the same places, which is time and space consuming task. In the Sloodle environment participants can be at remote locations and still attend training with their own pace.

In addition apparent advantages for participants, some features that are important from tutor's aspect have to be explained in more detail. First, connecting Moodle to Second Life through Sloodle is straightforward, and user documentation is well presented. Tutor himself can do it without additional help.

By using Sloodle tools WebIntercom and MetaGloos, tutor can significantly improve communication and collaboration among the participants. Prim Drop tool allows users' avatars to submit assignments in Second Life to the Moodle site. This tool provides the tutor with opportunity to enhance creativity of assignments.

Basics of Linden Scripting Language can be learned in a relatively short period of time. Having some previous programming knowledge could be useful, but not mandatory. Primary purpose of College *of Scripting, Music, and Science, Horsa Island* is to teach how to write LSL scripts. By applying the scripts, our 3D models can rotate, move, and change its color and shape.

From the researchers own experiences, there are two main drawbacks to consider when using Sloodle. Although basic features of virtual world are free to explore, and medical course can be temporarily placed on above mentioned island, having land for permanent course maintenance and importing graphical models are chargeable.

TABLE I.
Evaluation of VLE course

Question	*Satisfaction*
Access course and account	
The virtual environment allows linking orders with other platforms (Gmail, Facebook ...).	3
The virtual environment allows you to edit the user profile and enter additional information about yourself.	5
The virtual environment allows the user to view other users' accounts and insight into their current presence.	4
Structure of course and resources	
The environment allows the user to reorganize the section and change their schedules.	5
The environment allows tutor to turn on/off additional blocks (plug-in) that (do not) want to see in a course.	4
The environment allows the adaptation of the complete structure of the course according to tutor's needs.	5
Communication tools	
The virtual environment has tools for communication among participants.	5
If the system uses any form of synchronous communications, how useful are these components?	5
If conferencing software is integrated with the system in order to support group discussions and group working, how satisfactory is it?	4
Progress Tracking Tools	
How well does the tool allow tracking personal progress?	5
How rich a picture of an individual participant's interests and aspirations does the tool provide?	3
Technical Issues	
How easy is it to set up profile?	4
How easy is it to learn to use the system?	3
How time-consuming is it to enter data in profile?	4
How easy is creating graphical models?	2
How easy is working with LSL scripts?	4
Collaboration issues	
Does the tool provide shared workspaces? If so please rate the quality	4
How well does the VLE provide development of individual learning plans?	4
How easy is it for participant to collaborate in constructing a problem within the virtual environment?	4
How well does the system support collaborative working of a number of participants on the same project?	5
Does the system support submission of assignments from participant to tutor? If so please rate the quality	5
Does the system support recording and return of assessments to participants? How well does this facility meet tutors needs?	5

Second and more important drawback lies in complexity of vertebrae models. They cannot be easily created in Second Life. For their creation it must be used some CAD/CAM program, such as CATIA. And even if tutor knew how to use it, because of models complexity,

he would probably still need help of an expert in this field. Importing 3D models to Second Life can also be potential problem, but this issue is resolved by introducing program Blender. Nevertheless, this process can be time consuming.

V. CONCLUSION

Information learned in theory towards practical experience on real patients can be quite troublesome for patients as well as for students.

Although the majority of course development features are free, some still require payments, such as renting the land in Second Life, or creation of vertebrae models. Also, working with objects with demanding graphics requires additional skills that can be learned in reasonable period of time. Sloodle itself provides opportunities for upgrading, and self-developers can contribute.

Future work will be oriented towards overcoming of perceived drawbacks. Course introduction to medical staff users is also planned with further course detailed user evaluation.

ACKNOWLEDGMENT

The paper presented is supported by the Serbian Ministry of Education, Science and Technological Development (project III41007).

REFERENCES

[1] A. I. Albarrak, "E-learning in Medical Education and Blended Learning Approach", *Education in a Technological World: Communicating Current and Emerging Research and Technological Efforts*, December 2011, ISBN (13): 978-84-939843-3-5, pp. 147-153

[2] Modular Object-Oriented Dynamic Learning Environment. Available at: https://moodle.org/ last access 10.9.2014.

[3] Second Life. Available at: http://secondlife.com/ last access 10.9.2014.

[4] O. Yasar, T. Adiguzel "A working successor of learning management systems: SLOODLE", *Procedia Social and Behavioral Sciences*, Volume 2, January 2010, pp. 5682–5685

[5] Simulation Linked Object Oriented Dynamic Learning Environment. Available at: http://www.sloodle.org, last access 10.9.2014.

[6] G. Raymond, J. McKimm, "Medical Education and E-learning opportunities in the South Pacific", Samoa Medical Journal, Volume 3, Issue 3, December 2010, ISSN 2076-7994, pp. 32-41.

[7] B. Peck, C. Miller, "I think I can, I think I can, I think I can...I know I can Multi-user Virtual Environments (MUVEs) as a means of developing competence and confidence in undergraduate nursing students An Australian perspective", *Procedia Social and Behavioral Sciences 2* (2010), pp. 4571–4

[8] A. Uslucan, N.Senyer, "SLOODLE: Usage as an Educational Tool," *3dh World Conference on Innovation and Computer Sciences (INSODE 2013)*, vol. 04(2013), Antaliya, 23-28 April, 2013. pp. 745-752.

[9] F.B.Nunes et al, Integrating Virtual Worlds and Virtual Learning Environments through Sloodle: From theory to practice in a case of study for teaching of algorithms, *Nuevas Ideas en Informática Educativa TISE 2013*, 598-601

[10] Mitrović A., Milošević D., Božović M., Šendelj R., "Implementation of Computer Graphics course in Sloodle environment", International Electrotechnical and Computer Science Conference, ERK 2010, Portorož 2010.

[11] Maja Božović, Danijela Milošević, Marija Blagojević, Anđelija Mitrović, "Applying sloodle virtual environment for medical course preparation", *5th International conference E-learning*, Belgrade, 22-23 September, 2014. pp. 126-131. (ISBN: 978-86-89755-04-6)

[12] College of Scripting, Music, and Science, Horsa. Available at: http://maps.secondlife.com/secondlife/Horsa/184/249/2002

Massive Open Online Courses: edX vs Moodle MOOC

Marija Blagojević*, Danijela Milošević*
* Faculty of Technical Sciences Čačak, University of Kragujevac, Serbia
marija.blagojevic@ftn.kg.ac.rs
danijela.milosevic@ftn.kg.ac.rs

Abstract—**This research provides the overview of general possibilities of two massive open online courses edX and Moodle MOOC. The presented results show the comparative analysis of edX and MOOC pointing out different categories. The short discussion of performed analyses is also presented. The future work will be focused on the analysis of advanced features which both environments provide.**

I. INTRODUCTION

The term MOOC (Massive Open Online Course) coined during 2008 and it was related to an online course "connectivism and Connected Knowledge", designed by George Siemens and Stephen Downes. According to [1], MOOC integrates the advantages of social networking, a collection of open educational resources and the support of experts in a relevant field.

"Massive" refers to the number of the course's participants, as well as the capacities of the course in terms of allowing access to a large number of activities. George Siemens defined "massive" as: “Anything that is large enough that you can get subclusters of selforganized interests. Three hundred plus students could be one benchmark; another could be Robin Dunbar’s number of 150 people, which is the maximum after which the group starts to create smaller fractions.” [2, page 26].

"Open" usually refers to free access to individual courses, and sometimes it also applies to open or open content platform.

"Online" refers to MOOC access via the Internet.

"Course" means organizing content in a given time interval, from a subject area, which contains a set of resources with clearly defined goals and outcomes.

With the increased expansion MOOC imposes the choice of the platform for the creation of courses. This paper analyzes Moodle Mooc [3] and edX [4] platforms. Moodle LMS has been used for about thirteen years in the educational process. Taking into account the needs of teachers regarding examination and application of Moodle possibilities, a short and structured MOOC was created. It is called "Learn Moodle". The aim of this course is to introduce teachers with a way the students understand the activities, as well as with the features that Moodle provides.

EdX platform provides a huge number of courses in different fields in one place. edX courses are organized by prestigious universities, with the possibility of obtaining the certificate. With the necessary Internet connection, participants receive in one place learning materials, consultation with the teacher, and the possibility of evaluating the acquired knowledge.

There is a large number of studies dealing with the use of MOOC. A way of facilitating the integration of ICT tools in teaching through open environments is described in [5]. The authors plan the preparation of MOOCs together with the preparation of creative activities with the evaluation of the individual characteristics of the courses, the success rate, etc. The research in [5] demonstrates that the organization of communication activities during the course, highly affects course performance. Starting from this point of view, special attention was placed on the course activities that enable communication between teachers and students in the course, as well as among the participants themselves.

The researchers in [6] are engaged in open learning environments for their openness beyond existing learning management system. There are a large number of the research dealing with the possibilities of MOOCs.. Bearing in mind that Moodle is a free LMS system which is predominantly used at universities; it has been selected for the analysis. According to [7] the number of Moodle users is approx. 69,602,223. On the other hand, the last few years there is significant increase in using MOOCs. According to [8] edX aims to reach no fewer than one billion of users, and is used by the prominent institutions such as the Massachusetts Institute of Technology (MIT) and Harvard University.

II. PURPOSE, TASKS AND GOALS

Bearing in mind the expansion of MOOC the purpose of the research relates to the comparative analysis of the possibilities of MOOC Moodle and edX. The aim of the research was to determine the capabilities of both platforms in terms of teachers (course creators). Specific research tasks have been derived from the research objectives:

- Access to the two mentioned platforms (edX and Moodle MOOC)
- Creation of a course on the platform and edX and examination of designed course
- Examination of the possibilities of an open course within Moodle MOOC
- Comparative analysis of both platforms
- Overview and discussion of the capabilities for both platforms

III. METHODOLOGY

In order to create a comparative analysis of edX and Moodle MOOC, an account and a course have been created within edX platform in the field of programming language C. The platform is developed within the EU project Baeketel (TEMPUS project BEAKTEL "Blending academic and entrepreneurial knowledge in technology enhanced learning", http://baektel.eu). There is a number of courses within it, and some of them are shown in Figure 1.

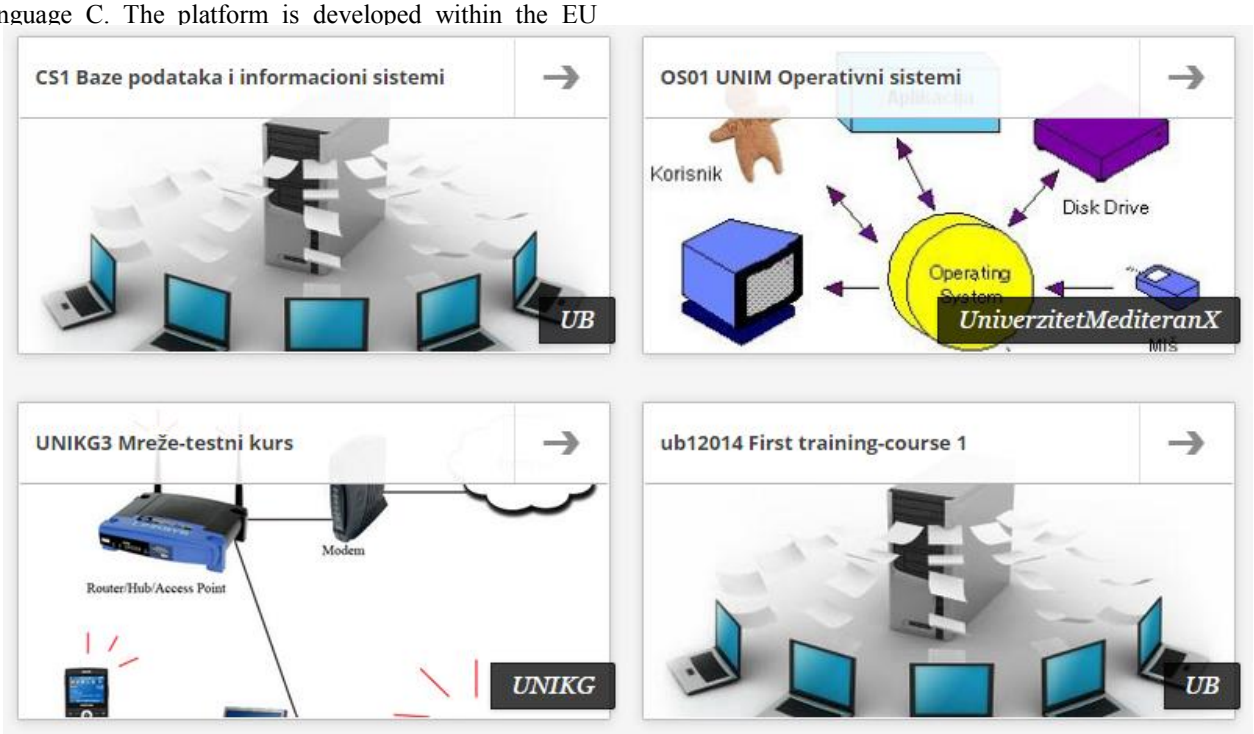


Figure 1. Overview of courses on edX platform

While the course analyzed within edX platform has been created, the same courses within Moodle MOOC were created already before. The comparative analysis of the features has been performed afterwards. Figure 2 gives the appearance of a part of the course within Moodle MOOC.

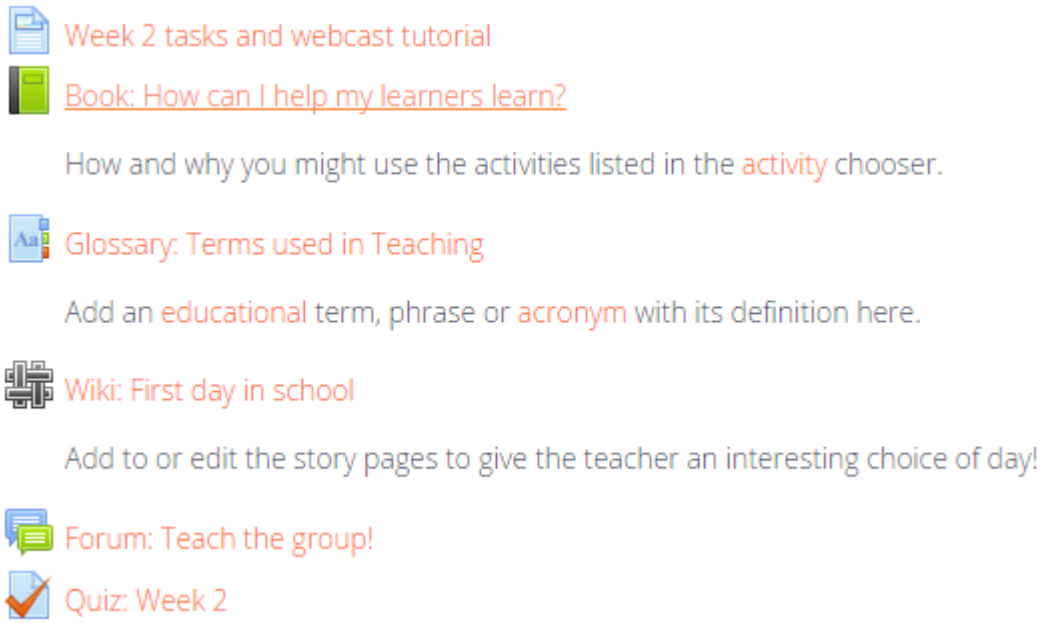


Figure 2. Overview of the course work within Moodle MOOC

IV. COMPARATIVE ANALYSIS OF EDX AND MOODLE FEATURES

Comparative review of the courses within Moodle MOOC and edX was done through following categories:

A. Course structure

In terms of course structure and organizational units the differences between edX and Moodle MOOC can be noticed. When accessing the topics and resources in Moodle MOOC, complete organization can immediately be perceived. List of topics with the contents are presented in the central part of the course, while the side plug-ins provide additional options.

Within edX structure list of content is located on the left side, while the content of the topic can be seen when you open a single topic.

Figure 3 shows a comparative overview of the structure of the course within Moodle MOOC and edX.

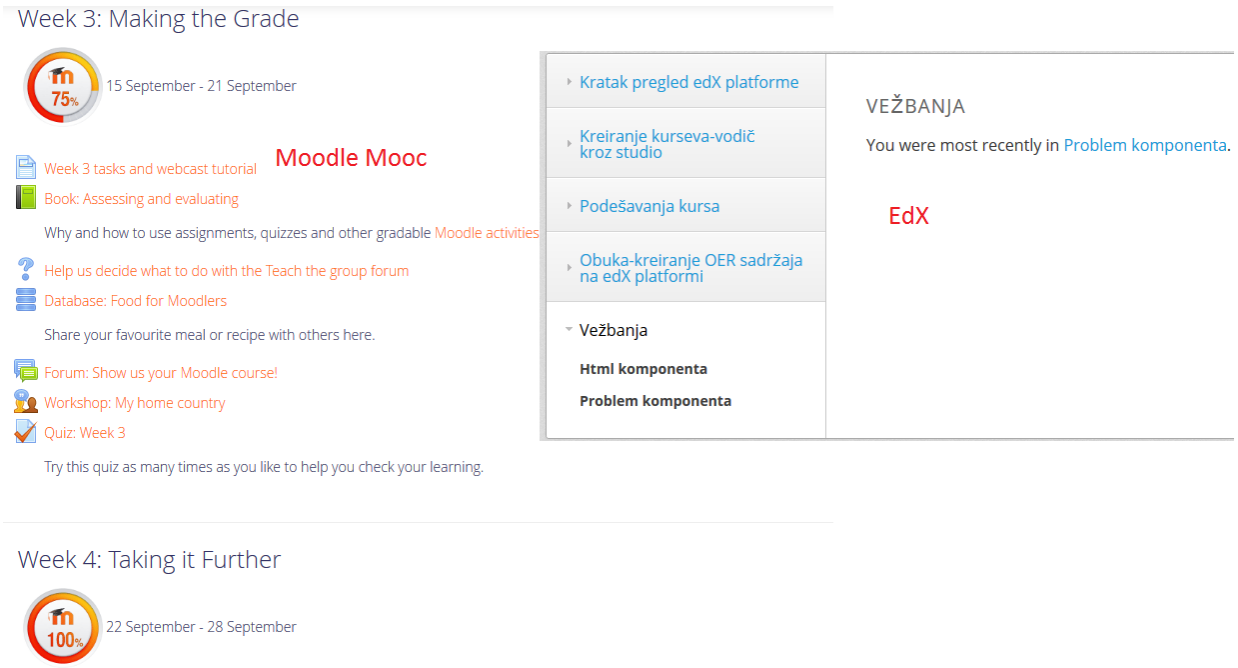


Figure 3. Comparative overview of the structure of the course within edX and Moodle MOOC

B. Collaborative modules

Moodle MOOC provides a possibility of organizing collaborative activities, through wikis and workshops. edX offers a possibility of collaborative activities for participants through the wiki, and also virtual labs are planned for more participants to work together.

C. Communication tools

Both systems posse the communication tools, in the form of a forum. However Moodle MOOC also has the possibility of communication through the chat rooms. Forums in both systems are organized so as to be used intuitively and without special preparation. Forums are meaningfully developed within Moodle MOOC because they provide more types of discussions.

D. Reports

Both systems provide the option of reporting on user activity. Moodle MOOC and edX provide an overview of the activities for each module that is placed in the course. Moodle MOOC allows detailed specification of activities, by user, user group, time in which the activity is taking place, etc. EdX also provides an overview of grouped activities. However, exports of these data provide more opportunities within Moodle MOOC, as compared to edX, in terms of possible formats in which the reports are exported. Visualization of the results is provided by both systems.

E. Tests and grades

Both systems provide an opportunity to test the participants, with the difference in the choice of the type of tasks, which is wider within Moodle MOOC. However, edX introduces the participants with advanced features setup tasks, such as mapping the correct answers in the pictures. In terms of defining score, edX provides the ability to define the scope of assessment, more meaningful regulation of the obligations of participants through the midterm and final exam, and it also provides the definition of the types of activities that will be included in the assessment.

Figure 4 shows a part of the possibilities of defining a grade in edX.

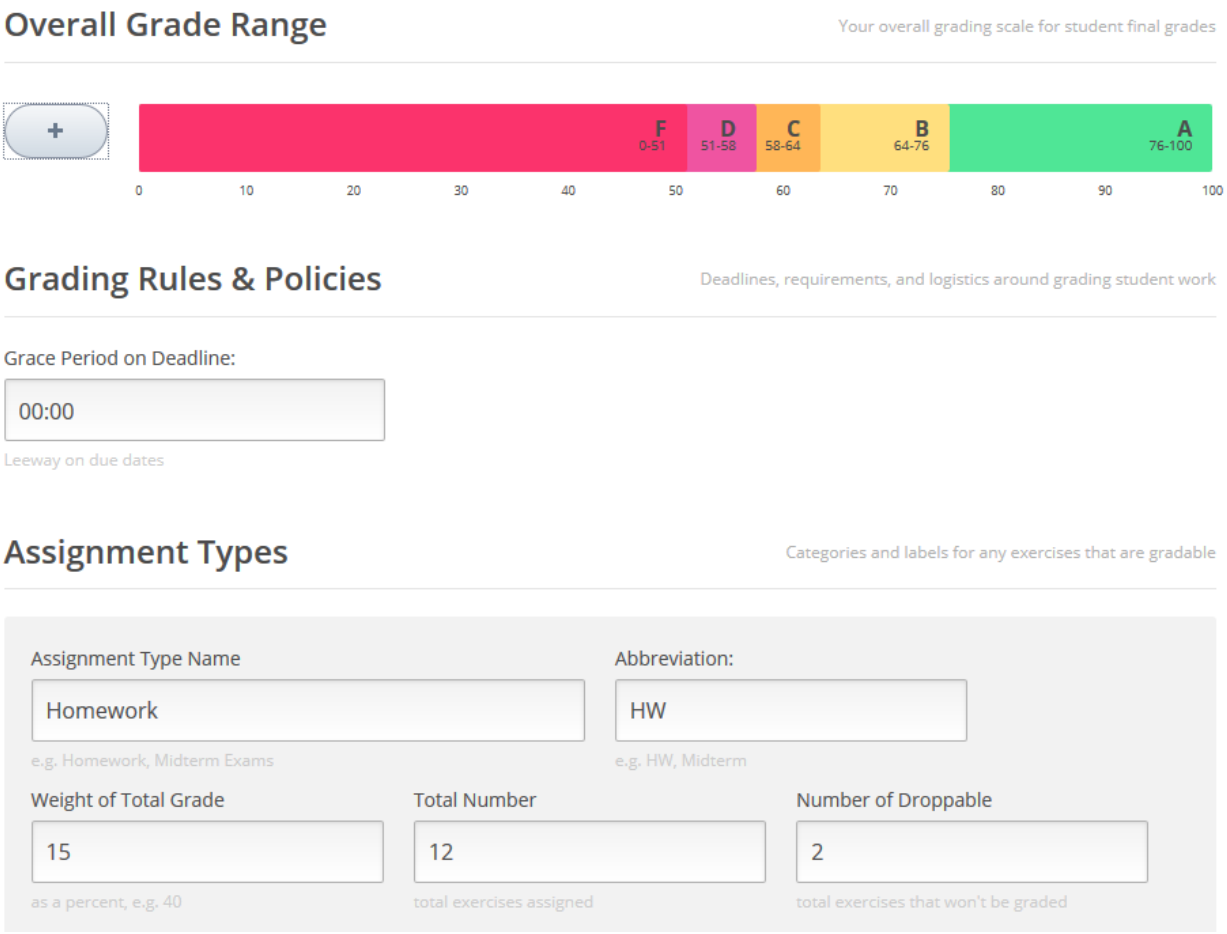


Figure 4. Overview of assessment opportunities within edX

V. EVALUATION AND DISCUSSION

The basic features of both systems that are commonly used in the classroom are listed in a comparative review of the items. The characteristics of both systems, with illustrative examples, are compared in more detail in [9]. This evaluation is the result of the subjective assessment by experienced user of both systems, unlike the one shown in [9]. The similar evaluation is performed in [10]. The evaluation of individual components, the appraisement of the available options on a scale from 1 to 5 (where 1 denotes the lowest satisfaction and 5 the highest), as well as the applicability of the remaining possibilities in teaching are made from the perspective of teachers is done. Table 1 shows the results of the subjective evaluation possibilities. As it can be noticed from the table, Moodle MOOC has advantages with communication tools, which is still under development within edX. On the other hand, edX has better assessment capabilities comparing to MOOC Moodle. Other categories are generally consistent with small differences in favor of one of analyzed systems.

TABLE I. SELF-ASSESSMENT OPPORTUNITIES WITHIN MOODLE MOOC AND EDX PLATFORM

	edX	Moodle MOOC
Course structure	4	5
Collaborative modules	4	4
Communication tools	3	5
Reporting	4	5
Tests	4	4
Grading	5	3

VI. CONCLUSION

On the basis of the obtained results, certain conclusions about the differences in the capabilities of EDX and Moodle MOOC can be drawn. Both systems support massive open online courses, but the individual segments differ regarding the features and use in terms of teachers. Depending on the specific needs the teacher in the classroom can make a choice between the analyzed platforms. The performed analysis has shown an original approach in the field of evaluation of both systems, while the similarities are reflected in the overview of its capabilities. Taking into account that according to [8] MOOCs may well be a leading element in the future of

higher education, the future work will be focused on the analysis of advanced features of both systems.

Acknowledgment

The work presented here was supported by the Serbian Ministry of Education and Science (project III 41007).

REFERENCES

[1] A. McAuley, B.Stewart, G.Siemens and D.Cormie, "Massive open online courses: Digital ways of knowing and learning", retrieved from: http://www.elearnspace.org/Articles/MOOC_Final.pdf, Last access: November 23 rd 2014.

[2] F.M. Hollands and D.Tirthali, "Moocs: Expectations and reality", retrieved from: http://www.academicpartnerships.com/sites/default/files/MOOCs_Expectations_and_Reality.pdf, Last access: November 23 rd 2014.

[3] Moodle MOOC, retrieved from:http://learn.moodle.net/, Last access: November 23 rd 2014.

[4] edX, Retrieved from: https://www.edx.org/, Last access: November 23 rd 2014.

[5] B.Lesjak and V.Florjancic, "Evaluation of MOOC: Hands on project or creative use of ICT in teaching", Retrieved from: http://www.toknowpress.net/ISBN/978-961-6914-09-3/papers/ML14-699.pdf, Last access: November 23 rd 2014.

[6] H. Fournier, R. Kop and G. Durand, "Challenges in research in MOOCs", Retrieved from: http://jolt.merlot.org/vol10no1/fournier_0314.pdf, Last access: November 23 rd 2014.

[7] Moodle LMS, statistics, Retrieved from: https://moodle.net/stats/, Last access: December 23 rd 2014.

[8] MOOCs: The Future of Higher Education? Retrieved from: http://www.topuniversities.com/student-info/distance-learning/moocs-future-higher-education, Last access: December 23 rd 2014.

[9] S. Kolukuluri, "edX-LMS Vs moodle-LMS and Performance Analysis in terms of number of users", Retrieved from: http://www.it.iitb.ac.in/frg/brainstorming/sites/default/files/Moodle_LMS__VS__edX_LMS.pdf, Last access: December 23 rd 2014.

[10] A Comparison of Five Free MOOC Platforms for Educators, Educators, Retrieved from: http://www.edtechmagazine.com/higher/article/2014/02/comparison-five-free-mooc-platforms-educators, Last access: December 23 rd 2014.

Adaptation of Online Courses for Students with Different Educational Backgrounds and Predispositions for Learning

Milena Frtunić*, Leonid Stoimenov*
* Faculty of Electrical Engineering, University of Niš, Niš, Serbia
milena.frtunic@elfak.ni.ac.rs, leonid.stoimenov@elfak.ni.ac.rs

Abstract— **In this paper a proposal for adaptation of course Algorithms and Programming will be presented. This course is in first year of academic studies at faculty of Electronic Engineering in Niš, University of Niš. In the past decade different approaches for adaptation in e-learning systems have been proposed. Also, categorization of students can be done in several ways and adaptation and personalization can be implemented for groups of students and for individual students. All this provides variety of possibilities for realization and it is important to choose the best approach for specific case. Within this paper, we will discuss and present the best solution for adaptation of the course which is attended by students with different educational backgrounds and predispositions for programming and algorithm creation, in order to facilitate the learning of the course material.**

I. INTRODUCTION

It is not a rare case that one course is attended by students with different backgrounds and predispositions for learning. This is especially common in the first year of studies on faculties where students choose departments in the second year of their studies. Educational program of such faculties is usually conceived to have at least one elementary course from each department of the faculty, on the first year of studies. In some cases, content of these courses can be quite challenging for students with different predispositions for learning to comprehend and learn.

At the Faculty of Electronic Engineering at University of Niš, all students must take the course Algorithms and Programming in their first year of studies. Within this course, they learn the basics of algorithm's creation and programming. Teachers' experiences from this course have shown that each year, there are two groups of students. One group consists of students that pass this exam without any problems in first examination term. In the other group are students that have a big problem with learning and passing the exam. Usually students from second group leave this exam as one of the last exams on their studies. Also, the experiences of the teachers show that students that pass this exam without any problems, later usually choose Computer Science Department to continue their studies and those that had problems passing the exam, choose other departments.

Based on the experiences of professors and teaching assistants on this course it was established that each year there is huge number of students that have problems with learning this course's materials and passing the exam. The main reason for this is that this course is attended by students with different background. They come from different high schools and some of them have never learned any basics that can be very helpful for learning this course's material. Those students have more challenges during preparing this exam. Another reason for this situation is that all students on Faculty of Electronic Engineering must take this exam. Many of them don't plan to go to Computer Science department and have no interest in programming and construction of algorithms. Furthermore, many of them don't have predispositions for this type of subject. Because of that, for them learning this material can be very difficult and thinking in that way can be very challenging.

Because of these differences between the students, it is important to use possibilities that come with e-learning systems and adapt and personalize the content of the course, so that all students can learn it more easily and pass the exam without too many difficulties. Also, it is extremely important that the content of the course remain the same, since it contains only the basic knowledge of the subject which all students that attend the Faculty of Electronic Engineering should be familiar with.

In order to achieve this goal, the adaptation and personalization of the course's content based on the student's background and predisposition for learning has to be done.

In the next part of the paper, adaptation and personalization in e-learning will be discussed first, and special attention will be given to approaches for adaptation of lectures. Also, adaptive e-learning systems will be explained with focus on ways for providing adaptation in e-education system. Further in the paper, the proposal for adaptation of course Algorithms and Programming at the Faculty of Electronic Engineering at University of Niš, will be given. Within this part of the paper, categorization of the students will be discussed, along with methods of their categorization into groups.

II. ADAPTATION AND PERSONALIZATION

With new possibilities in the world of teaching and learning in the past decade, personalization and adaptation of the e-learning systems is becoming important characteristic of these systems.

When designing system's adaptation it is very important to understand student's differences. Student diversity can be observed through three aspects [1]:

- Learning style – this concept refers to the fact that some students prefer specifics and observable phenomena and other are more comfortable with theories and abstractions. This concept also refers to the fact that some students better understand and remember visual information and others make better progress with verbal explanations [2].
- Approaches to Learning and Orientations to Studying – students have different approaches to learning. Some students have surface approach and remember only most important things and make no effort to understand the point of the lesson. Others have deep approach and go into details with every subject and finally there are those that have strategic approach and learn only what is needed to get the highest grade.
- Intellectual Development – most of the student think that knowledge is certain and they only need to learn same facts and repeat them at some point. But there are some students that see that knowledge is contextual and that they should make their own conclusions based on evidence and not only on professor's word.

For all these groups different approaches are required in order to personalize e-learning system so that it can be suitable for everyone. However, in the past decade term personalization of these systems is referring to [3]:

- personalization of the learning content based on student's preferences and background,
- personalization of representation of the learning content,
- and combination of the two.

Adaptation of e-learning systems refers to a process where learning content is being delivered to learners adaptively, which means that the appropriate contents are delivered to the learners in an appropriate way at an appropriate time based on the learners' needs, knowledge, preferences and other characteristics [4]. It can be done by one of these criterion or few of them, based on the goal that is set to be achieved.

Adaptation and personalization of e-learning systems can be done for a group of students and for each student individually. Adaptation for each student individually means that every student in the system will have his own personal space which will be adapted based on his habits, interests, preferences and wishes. It can be done in different levels, from adapting only user interface to adapting every aspect of it with focus on the course's content. Since this can be very challenging, most systems today make adaptation for groups and not for every student individually. First they form a certain number of groups which are combination of criteria they want to use for adaptation of e-learning system and then they split all students in those groups by assigning each student to one group.

Personalization of LMS nowadays usually has focus on controlling which courses the student is allowed to view and restrictions within the course. Some e-learning systems allow personalization of student's profile, which refers to creating personalized calendar, editing user interface and adding additional widgets. These solutions usually do not include any kind of automatic adaptation of learning content based on student's preferences and learning style.

Recently, there has been a progress in creating systems that adjust content based on student's profile, his habits, learning style, skills and preferences. There are many different ways this branch can be studied, developed and where the focus can be put. One solution that includes some of the characteristics is presented in [5] and refers to adjusting content by using environmental and location information for mobile learning in field of environmental sciences.

In the field of learning foreign languages, there were attempts of development of systems that recommend reading material based on knowledge level and preferences, and reading annotations which can be individual and shared [6].

Huge attention has been given to the e-learning systems that are intended for people with special needs for learning. These systems have focus set on the special needs of the users and adaptation and personalization is done based on those needs [7] [8].

On the other hand there are systems created for general use and not for some special area which provide adaptation of learning content. Some scientists went in another direction and proposed multi-agent systems that are learner-centric and improves learning outcome, satisfaction of learners and enhances education [9]. This approach uses multi-agents which communicate with each other and create content for each student based on his knowledge level and skills (past), preferences (present), learning performance, and objectives (future).

Most of all systems that provide personalization of content use algorithms and methods that choose proper content and subject based on user's profile. These systems make decisions regarding what would be interesting for the user, choose topics that are similar to those he already know, define a level of knowledge he can gain from the course based on his predispositions and present the content in personalized way. But for these systems it is not common to have predefined knowledge that one user should learn and do the adaptation of that system so that user can be comfortable learning it.

III. COURSE ADAPTATION

Within this paper we will discuss the best approach for adaptation and personalization of the course Algorithms and Programming at the Faculty of Electronic Engineering at University of Niš, in order to provide better material and learning conditions for different types of students. First, we will discuss the best way for dividing students into groups for which the adaptation will be done. After that, description of the workflow will be provided and at the end, solutions for personalization and adaptation will be proposed. Since the Faculty of Electronic Engineering in Niš uses Moodle LMS [10], we will propose a solution that uses possibilities of Moodle LMS in order to make adaptation for lessons' content.

A. Student types

In the previous part of the paper, it was explained that adaptation of e-learning systems and material can be done for groups of students and for individuals. Regardless of the choice, it is important to establish criteria by which the adaptation will be made. As it was already presented in

the paper there are different ways for establishing criteria for adaptation. In the case of the Algorithms and Programming course, those methods and algorithms are not suitable, since the goal is to teach all students the same material. The aim is only to adapt the presentation of the material based on the group they belong in. Because of that, we have decided to group students by two characteristics that we find relevant for adaptation we want to provide:

- by the background they have at the beginning of the semester,
- by their predispositions for programming and interest in the subject.

1) Gouping students based on their background

A first criterion for dividing students into groups is based on their educational background. This criterion is chosen because not all students come from high schools with same educational program. This results in having students that had never had any subjects that would give them basics for this course. On the other side, there are students that have done some programming in high school and have more than enough educational background to attend the course. These differences between the students' background require different approaches while explaining material to those students with no previous knowledge and to those that have some. Fig. 1 presents groups defined by student's educational background at the beginning of the semester.

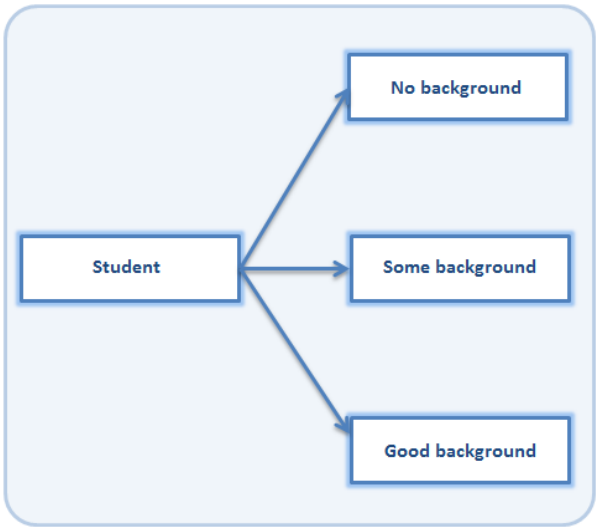


Figure 1. Groups of students based on their background

As it can be seen from the Fig. 1, all students are split into three groups based on their previous knowledge:

- Students with no educational background – this group consists of those students that have absolutely no educational background from this field of subjects. They come from high schools where they did not attend any subject from this field and are expected to have some problems attending the course and following the lessons. These students require more attention during introductory lessons of the course, in order to get them closer to the subject and enable them to more easily follow the classes later in the semester.
- Students with some educational background – this group consists of students that have enough educational background to attend the course with no problems in regards to understanding the subject and attending the lessons.
- Students with good background – this group consists of students that have more than enough educational background to attend this course. They come from special high schools for Computer Science and Informatics and have done enough programming during the high school.

2) Gouping students based on their predispositions for programming and algorithm creation

A second criterion for dividing students into groups is based on their predispositions and interest in the subject. Decision to include this criterion is based on the experience of the teachers involved with the course. The experience showed that many students that have attended this course do not have predisposition for programming and thinking in the way that is natural for programmers, regardless of previous experience. Because of that, this course is very hard for them and they have many problems in passing the exam. For that reason we decided to put this as a criterion for grouping students. Fig. 2 presents defined groups based on this criterion.

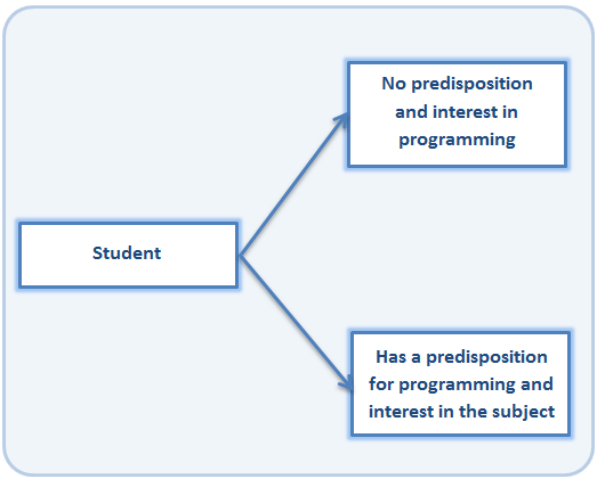


Figure 2. Groups of students based on their predispositions and interest

As it can be seen from Fig. 2, all students are split into two groups based on their predispositions for the subject:

- Students with no predispositions for the subjects – are students that have no predispositions and talent for programming and constructing algorithms. Usually these students entered the Faculty in order to choose some other department and not the department of Computer Science. Since they have no talent for this kind of subject, they need special attention in order to learn the teaching material of this course. It is important to teach those students to think in a way needed for this subject and to approach lessons in a way that will be is familiar to them.
- Students with predispositions for the subject – are students that probably won't have any problems with this course. They have good predisposition for the course and are interested in programming and

algorithm construction. These students already know how to think in that way and needn't any special approaches for teaching. These students entered this Faculty with a goal to continue their studies on Department of Computer Science.

B. *Student categorization*

Students should be split into groups at the beginning of the semester. Fig. 3 presents a flow for categorizing a student into appropriate group.

Since Faculty of Electronic Engineering in Niš uses Moodle LMS [11], each student has a profile at Moodle LMS and at the beginning of the semester they enroll in the course. As it can be seen from the Fig. 3, at the beginning of the flow, student logs on the Moodle system and enrolls in the course.

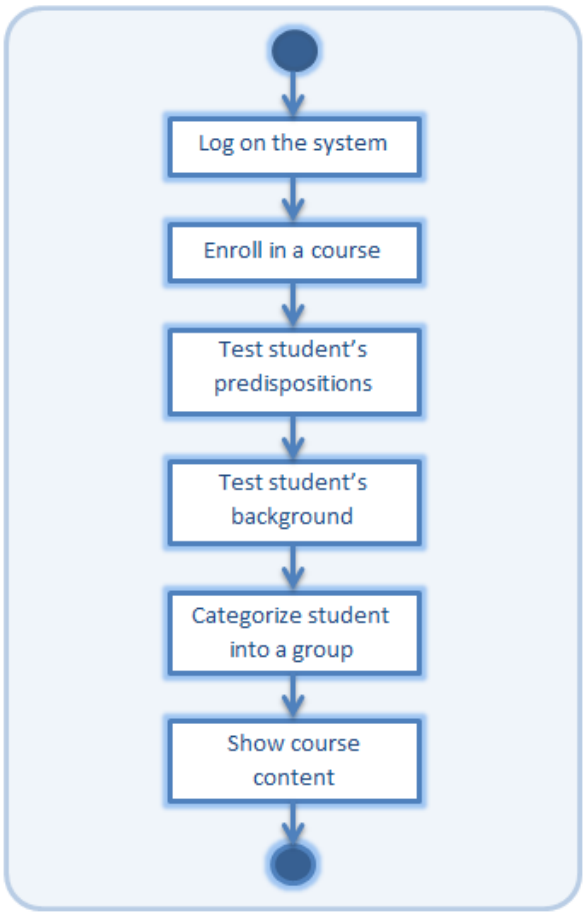


Figure 3. Categorization of students into groups based on their background, predispositions and interests

At this point students should be divided into groups based on criterion that was presented in previous part of the paper. This is done with two-step student testing: first step is to determine their predispositions and after that to determine their background. The results of both tests will show in which group student belongs. Testing can be done by using appropriate Moodle Quizzes. For this part, mobile application MoodleQuiz [12] can be a very useful tool for testing students, since it provides possibility to test them out of the classrooms.

First, student will take the test that estimates their predispositions for the subject. Questions for this test will be constructed in consultation with psychologist in order to make reliable estimation of student's predispositions. At the end of the test student won't be able to see the test results, since test is not of such nature.

In the second step, students will take the test that estimates their educational background. Questions for this test will be prepared by teachers, since they can make the best estimation of the previous knowledge that is enough for student to take the course. Correct answers in the quiz will not be presented back to the students, since that is not relevant in this case.

After both tests are done, system will categorize students by their results into groups. At the end, students will receive only a number of the group they were categorized in and beside the regular content of the course, adapted content for their group will be presented to them.

Although two criterions that are used for student's categorization make six combinations, we will create four groups of students (four student's profiles). Fig. 4 presents groups that are created based on both criterions. As it can be seen from Fig. 4, groups are defined as: profile 1, profile 2, profile 3 and profile 4.

In *profile 1* are students that have no educational background and have no predispositions for this kind of subject.

In *profile 2* are students that have no educational background but have predispositions for this type of subject.

In *profile 3* are students that have some educational background or have good educational background and have no predispositions or interest in the subject.

In *profile 4* are student that have some educational background or have good educational background and have predispositions and interest in the subject.

Decision to group students with some background and those with good background in the same profile was made because students from both groups are able to follow lessons without problems and does not need special approach for explaining the lessons.

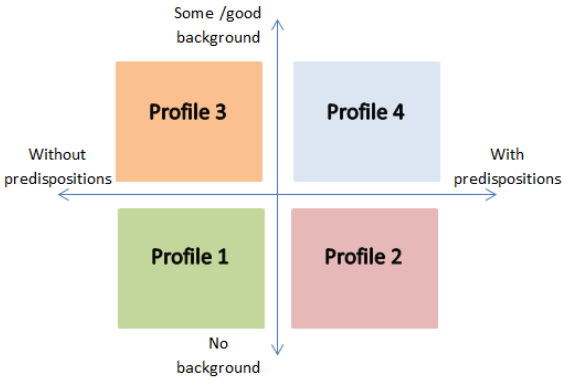


Figure 4. Student's profiles

C. *Course adaptation*

After students are categorized into groups and their profiles are completed, they will be offered two sets of course material:

- standard course material, that is same for all students.

- personalized material that is adapted for group they are assigned to.

In that way all students will be able to see standard lectures, but also they will have possibility to learn from adapted material. The content of both materials will cover all topics but the approach in explanations will be different in adapted material.

In this model, personalization of course's content and its adaptation is done by teachers on the course. At this point, it is considered that this should be done manually and that they are most adequate for this task.

For presentation of personalized course's content Moodle Lessons [13] are used. The idea is to create four Moodle lessons for each topic on the course. Each lesson will belong to one group of students and will be adapted according to their needs. Moodle Lessons module provides many possibilities for setting up the lesson [14]. For purpose of adaptation and personalization of the lessons we believe that the following settings are appropriate:

- Lessons will be password protected and student will know passwords only for lessons that belong to his group.
- Each lesson will contain questions that will navigate the student through the lesson.
- Student will be able to exit the lesson whenever he wants, but will be obligated to start from the beginning the next time he opens the lesson.
- Student can open the lesson as many times as he wants.
- Lessons will be used only for learning and practice so student won't be graded by success on them.
- Lessons will depend on each other, so that student will have to complete previous lessons if he wants to open next. In this way we are ensuring that student have background knowledge for the lesson he is opening.
- Lesson will be available after the lessons topic is presented by the teacher in the classroom.
- Lesson will not have a deadline.

With these setting we will provide lessons that will guide students through the topic and evaluate their knowledge at the same time. After every lesson they will know how well they have learned the lesson and if there is something they did not understand.

Since this module provides easy way of customization of content, it is a suitable tool for presenting material in different ways, appropriate for different groups of students.

This model for adaptation and personalization of the course will be supported by one service presented in Figure 5. After student finishes the test of his predispositions and the test of his educational background, this service will process answers that student gave on the test and put student into appropriate group. After that, service will store this information into Moodle database and send group's number to the student. Service will be in charge for retrieving and using this information when it is required. This will be done when lesson's password is sent to the students. This service will send only passwords for those lessons that are within the group that student belong in. Furthermore, this service will be responsible for presenting appropriate lessons to the students, so that student sees only his group's lessons and not all of them.

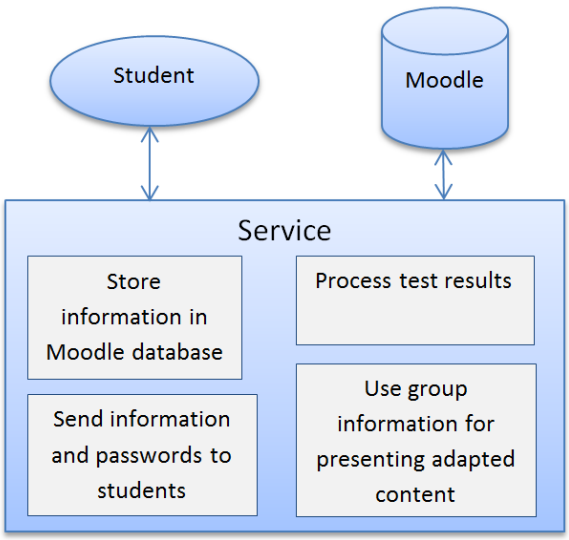


Figure 5. Service for handling student's groups in Moodle LMS

IV. CONCLUSON AND FURTHER WORK

In this paper we have described adaptation of the course Algorithms and Programming at the Faculty of Electronic Engineering in Niš. But this model of adaptation is suitable for any course that requires fixed knowledge of all students by the end of the semester and is attended by students with different educational background and predispositions. The proposal for presenting the personalized material is done by using Moodle LMS, since it is used on Faculty of Electronic Engineering in Niš.

In order for this model to become functional, some adaptations of Moodle Quiz are required, in this case that refers to MoodleQuiz application. The adaptation refers to the last part of the application, when quiz results are presented. Since this testing does not require results to be presented back to the student, handling this type of quizzes should be added to the application.

Furthermore, this system requires a development of a service that will handle student groups for adaptation of course's content, which includes processing test results, storing and retrieving group information and using it for course adaptation and personalization.

Also, this model includes small adaptation of Moodle and Moodle database so that information about a group to which a student has been assigned in, can be stored in the database and properly used when needed. This means expansion of Moodle database and implementation of functions that will cover this functionality.

As it is presented in previous part of this paper, with this model for teaching, at this point we offer students both personalized and non-personalized content. Our goal at this stage is to create prototype of this system and test it. First goal is to put student into appropriate groups, so that material is suitable for them.

If the prototype proves successful in dividing students into groups and if the statistics show that exams passage rate grows and students find it acceptable to use, we will

move to second stage, which means cutting non-personalized material content out of the course.

ACKNOWLEDGMENT

The research presented in this paper was funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia as part of the project "Infrastructure for electronically supported learning in Serbia" number III47003.

REFERENCES

[1] R. M. Felder, R. Brent, "Understanding Student Differences", *Journal of Engineering Education*, 94 (1), pp. 57-72, January 2005

[2] R. M. Felder, L. K. Silverman, "Learning and Teaching Styles In Engineering Education", *Engineering Education*, 78(7), pp. 674-681, 1988

[3] R. Pavlov, D. Paneva, "Personalized and adaptive elearning – Approaches and solutions"

[4] L. Shi, Alexandra I. Cristea, J, G. K. Foss, D, Al Qudah, A, Qaffas, "A Social Personalized Adaptive E-Learning Environment: A Case study in Topolor", *IADIS International Journal,* pp. 01-17, 2013

[5] P. Seyedabolghasem, A. K. Haghi, K. Morovati, "Presenting a personalized mobile learning recommender system by using environmental and location information", *IOSR Journal of Engineering (IOSRJEN),* Vol. 3, Issue 3, pp. 01-09, Mart 2013

[6] H. Ching-Kun, H. Gwo-Jen, C. Chih-Kai, "A personalized recommendation-based mobile learning approach to improving the reading performance of EFL students", *Computers & Education,* Volume 63, pp. 327–336, April 2013

[7] Á. Fernández-López, M. J. Rodríguez-Fórtiz, M. L. Rodríguez-Almendros, M. J. Martínez-Segura, "Mobile learning technology based on iOS devices to support students with special education needs", *Computers & Education*, 61, pp. 77–90, 2013

[8] TERENCE, [Online], Available: http://www.terenceproject.eu/ [Accessed 20.12.2014].

[9] K. N. ElSayed, "Individual Syllabus for Personalized Learner-Centric E-Courses in E-Learning and M-Learning", *(IJACSA) International Journal of Advanced Computer Science and Applications,* Vol. 5, No. 6, 2014

[10] A. Stanimirović, L. Stoimenov, "Implementation of blended learning environment using Moodle platform," in *The Second International Conference on e-Learning (eLearning 2011)*, Belgrade, 2011, pp. 163-168

[11] "Moodle LMS," Moodle, [Online]. Available: https://moodle.org/. [Accessed 20.12.2014].

[12] M. Frtunić, M. Bogdanović, and L. Stoimenov, "Moodle quiz on android mobile devices," in *YUINFO 2014*, Kopaonik, 2014, pp. 07-12

[13] "Moodle lessons," Moodle, [Online]. Available: https://docs.moodle.org/28/en/Lesson_module [Accessed 20.12.2014].

[14] "Moodle lesson settings," Moodle, [Online]. Available: https://docs.moodle.org/28/en/Lesson_settings [Accessed 20.12.2014].

Multi linked lists: an object-oriented approach

Đorđe Stojisavljević*, Eleonora Brtka**, Vladimir Brtka**, Ivana Berković**
* University of Banja Luka/Faculty of law, Banja Luka, Republika Srpska
** University of Novi Sad/Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia
djordje.pfm@gmail.com, eleonorabrtka@gmail.com, brtkav@gmail.com, berkovic@tfzr.uns.ac.rs

***Abstract* — The paper deals with the approach to multi linked lists while teaching. Object oriented paradigm is used, so that multi linked lists are implemented in C++. Simple example presented in this paper cover the usage of structures inside classes and template classes. Basic concepts of understanding object oriented multi linked lists are defined, and five groups of students are singled out. The main contribution of this research is in the domain of education: the specification of student's understanding of these concepts is given, as well as guidelines how to recover to full understanding of multi linked lists.**

I. INTRODUCTION

Linked lists are widely known and exhaustively described in literature. An elaborate discussion of linked lists can be found in e.g. [1], while more detailed discussion about multi linked lists and their implementation in C language can be found in e.g. [2].

Each node of a multi linked list (Fig. 1) has a complex structure; it contains:

- Data field – represents the useful data, usually realized in structure form;
- One link field that points to the next node in the multi linked list (like in singly linked list). The last node points to NULL; and
- Two or more link fields who are pointing to another lists called *sublists*. Sublist has a structure like singly linked list.

The entry point into a linked list is called the *head* of the list. It should be noted that head is not a separate node, but the reference to the first node. If the list is empty then the head is a NULL reference.

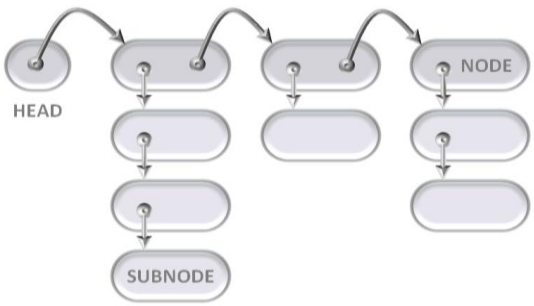


Figure 1. Multi linked list in traditional (structured) form

As we can see from Fig. 1 there are two structures:

- *subnode* which contains data field and link to the next subnode, and
- *node* which contains data field, link to the head of sublist and link to the next node.

These concepts are often hard to understand and implement in practice. Object Oriented (OO) programming paradigm is most common contemporary programming paradigm, so it is necessary to implement multi linked list in OO manner. The OO approach to multi linked list is even more confusing if not presented properly to the students. So, main questions are

- How to deal with OO C++ multi-linked list while teaching?
- What are the basic concepts?
- What to do with students who do not understand some of basic concepts?

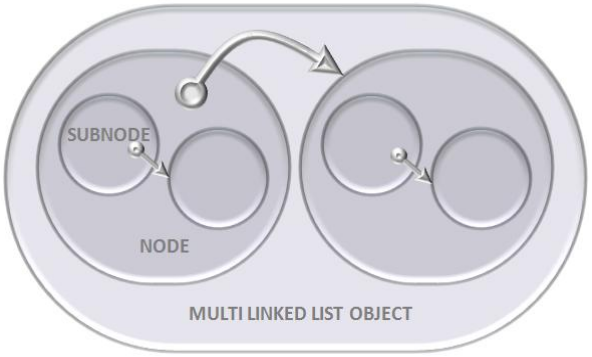


Figure 2. Multi linked list in object-oriented form

This paper is organized as follows: Section II deals with OO approach to multi linked lists. Simple example in C++ programming language is used to declare multi linked list, as well as constructor method, destructor method and some operations on multi linked lists. Section III presents the methodological approach to multi linked lists; six basic concepts were defined, while students were classified to five distinctive groups according to their understanding of six basic multi linked list concepts. There is no point to consider students who understand all basic concept in full extents, so neither of these five group covers them. Section IV is the conclusion of this research where some guidelines on how to recover to fully understanding of OO multi linked lists are given, for each group of students.

II. OBJECT – ORIENTED APPROACH

A definition of object-orientation is that an entity of whatever complexity and structure can be represented by exactly one object [3]. If we apply this definition on multi linked list we get an object-oriented multi linked list (Fig. 2). In object-oriented programming we treat a multi linked list as an object. That means that we will view a multi linked list as an object that stores data as a list, that allows

the list to be manipulated using a set of methods provided by the multi linked list interface. An important element to good object-oriented programming is good object-oriented design. This means that we need to design a good interface for a multi linked list object that provides the operations that a programmer wants. Design of a multi linked list is shown on UML class diagram in Fig. 3.

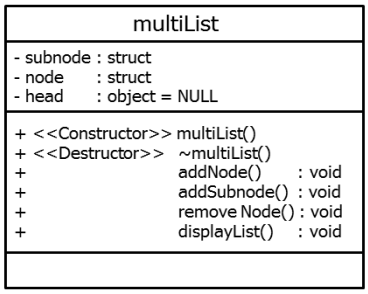


Figure 3. Multi linked list class diagram

As we can see, internal structure of a list is hidden. List can be manipulated only through the interface.

A. Multi linked list class

Implementation of a multi linked list will be hidden so that it can be modified without affecting the programs that use it. In particular, we will not let the programs that use it, to have access to its internal representation. Therefore, we will declare a multi linked list in C++ as shown in Listing 1.

```
template <class T>
class multiList{

private:
  struct subnode{
    T value;
    subnode *next_sub;
    subnode(T value1,subnode
*next_sub1=NULL){
       value=value1;
       next_sub=next_sub1; } };
  struct node{
    T value;
    subnode *head_sub;
    node *next;
    node(T value1, subnode *head_sub1=NULL,
         node *next1 = NULL){
           value = value1;
           head_sub=head_sub1;
           next = next1; } };
node *head;

public:
  multiList() { head = NULL; }
  ~multiList();
  void addNode(T value);
  void addSubnode(T n, T value);
  void removeNode(T value);
  void displayList();
  };
```

Listing 1. Declaration of a multi linked list in C++

As we can see from Listing 1. multiList class is realized as template class, because the class should be able to handle different types of data fields. While using a multi linked list and operating on a particular data type, only the data type needs to be specified when the template class object is defined or declared, e.g. `multiList<int> mList`.

B. Constructor and destructor

MultiList class has one constructor and destructor. Constructor *multiList()* has no arguments. Its function is to initialize multiList object by setting head to NULL.

Destructor gets called when multiList object needs to be deleted. Through while loop destructor runs-cross each node of a multiList object and deletes it by calling *removeNode()* method.

C. Methods

In order to make the multiList as universally usable as possible, we want to define a set of essential, primitive operations that programmers can use to assemble more complex operations. In other words, rather than trying to imagine every conceivable use for a multiList and placing an operation in the multiList's interface that supports that use, we try to envision a set of basic building block operations that can be used to create these more complicated operations.

To add a new node in the list, first thing to do is allocate memory space for the node, then we assign data to data field [4]. When node is created, it has no subnodes; therefore head of sublist is NULL. Next, we concatenate the new node with the original list and make the newly created node the first one of the list (Fig. 4).

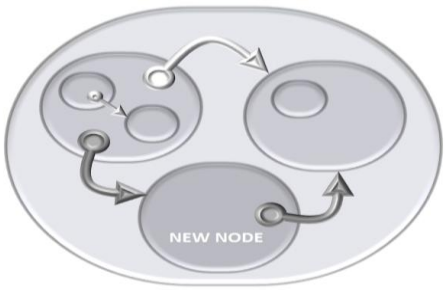


Figure 4. Adding a new node into the multiList object

Method for adding a subnode is given in Listing 2.

```
template <class T>
void multiList<T>::addSubnode(T n, T value){
   node *nodePtr=head;
        // check if node exists
        // if exists then add subnode
   if(nodePtr->head_sub==NULL)
   nodePtr->head_sub=new subnode(value);
   else {
      subnode *subnodePtr=nodePtr->head_sub;
      while(subnodePtr->next_sub!=NULL)
      subnodePtr=subnodePtr->next_sub;
      subnodePtr->next_sub=new subnode(value);
         }
}
```

Listing 2. Method that adds a new subnode into the multiList object

Removing a node from a multiList means to modify the list in such a way that the node is no longer connected to its predecessor and successor, while bridging the removed node to maintain the connection of the other nodes. After bridging, programmer explicitly frees the memory occupied by those nodes that are not needed anymore [2].

Method that removes the first element of a list is given in Listing 3.

```
template <class T>
void multiList<T>::removeNode(T value){
   node * nodePtr; node *previousNodePtr;
   if (!head)  return;
   // find the node that needs to be removed
   while(subnodePtr!=NULL){
   // remove his subnodes
        }
    head = head->next;
    delete nodePtr; }
    else {
        nodePtr = head;
       // update links
         }
    if (nodePtr){
          previousNodePtr->next=nodePtr->next;
           delete nodePtr; }
}
```

Listing 3. Method that removes a node from the multiList object

To perform an operation on all nodes of a list, we have to reach each node starting from the first one, by following the *next* references. The simplest way of doing this is through iteration [5].

In Listing 4. we give an example of using object-oriented multi linked list. Because the multiList class is a template class, in our example we will define it as a string.

```
int main(){
   multiList<string> list;
    string name;
    string value;

    cout << "Add 3 names to the List:\n";
    for (int i = 0; i < 3; i++){
       cout << "Name #" << i + 1 << ": ";
       getline(cin, name);
       list.addNode(name);
   }

   cout<<"Add subnode to name: ";
    getline(cin,name);
    cout<<"\nEnter value: ";
    getline(cin,value);
    list.addSubnode(name,value);
    list.displayList();
   cout<<"\nEnter a name to delete: ";
    getline(cin, name);
    list.removeNode(name);
    list.displayList();
   list.~multiList();
  return 0;
}
```

Listing 4. Test example of using multiList class

III. METHODOLOGICAL APPROACH

Methodological approach used to explain multi linked lists in the case when object oriented programming is applied is based on six concepts. These are basic concepts, arguably minimal number of basic concepts needed to practically understand object oriented multi linked lists. Basic concepts are:

1. Pointers and memory allocation.
2. Object-oriented paradigm.
3. Linked list basics.
4. Creating nodes.
5. Creating sub-nodes.
6. List operations.

The importance level of basic concepts is crucial for students to understand multi linked lists.

(The order of basic concepts is crucial for students to understand multi linked lists. These concepts were chosen after extensive research of literature references. In [6] was presented a "pointer-safe" object oriented paradigm including physical addresses, placements of objects, etc. in addition, in [7] the context-insensitive pointer analysis was described; this is based on applying cycle elimination to context-sensitive pointer analysis and refers to some advanced techniques. Object oriented paradigm is widely used in practice, so that there is no lack of literature references to this concept; in [3] this paradigm is described appropriately for this research. The linked lists concepts including basics, creating nodes and sub-nodes, as well as operations on lists are presented in [1, 4, 5, 8], and there is no lack of literature on this matter as well.

In contrast to a large number of literature references dealing with these concepts in the domain of software engineering, there is a lack of information about implementing these concepts in teaching. It is hard to assess the understanding of some concept, so we are not particularly sure if student understand these concept, and even less are we able to objectively assess the extent to which student understands a particular concept. The application of the scale (e.g. from 5 to 10 or from 1 to 10) is often used, as well as the measure of understanding in percents, but arguably more "rough" scale is better, so we are using just three values in this investigation: low (0), medium (1) and high (2). Instead of three-point scale, five or seven point scale is often used.

Still, there are some disagreements about basic concepts, so we had applied Fuzzy Screening method (R. Yager) and the Rough Sets Theory (Z. Pawlak). In both cases, we needed a data sample.

In this particular investigation we used data sample collected from multiple sources in mid-term exams. We have students, and each of them have a certain number of points ranging from 55 to 100 for each of this six attributes. Having in mind that this data sample is small and gathered from multiple sources we have discretized our data so that we have three values: low, medium and high: from 55 points to 70 points is low, from 76 to 85 is medium and from 86 points up to one hundred points is high. After discretization step, each row represents one or more students, Table I.

TABLE I
DATA SAMPLE

1. Pointers and memory allocation.	2. Object-oriented paradigm.	3. Linked list basics.	4. Creating nodes.	5. Creating sub-nodes.	6. List operations.
High (2)	High (2)	High (2)	High (2)	High (2)	High (2)
High (2)	High (2)	High (2)	High (2)	High (2)	High (2)
High (2)	High (2)	Medium (1)	Medium (1)	High (2)	High (2)
Medium (1)	High (2)	High (2)	Medium (1)	High (2)	High (2)
Medium (1)	High (2)	High (2)	High (2)	Medium (1)	High (2)
…	…	…	…	…	…
High (2)	Medium (1)	High (2)	Medium (1)	Medium (1)	Medium (1)
High (2)	Low (0)	Medium (1)	High (2)	High (2)	Medium (1)
Medium (1)	Medium (1)	Medium (1)	High (2)	Medium (1)	High (2)
Medium (1)	High (2)	Medium (1)	High (2)	Medium (1)	low (0)
High (2)	Medium (1)	Medium (1)	Medium (1)	Low (0)	Medium (1)
High (2)	Medium (1)	Medium (1)	Low (0)	Low (0)	Medium (1)
Medium (1)	Low (0)	Low (0)	Medium (1)	High (2)	Medium (1)
High (2)	Medium (1)	Low (0)	Medium (1)	Low (0)	Medium (1)
Low (0)	Medium (1)	High (2)	Low (0)	Medium (1)	Medium (1)
High (2)	High (2)	Low (0)	Low (0)	Low (0)	Medium (1)
Medium (1)	High (2)	Medium (1)	Low (0)	Low (0)	Medium (1)
Low (0)	High (2)	Medium (1)	Low (0)	Medium (1)	Low (0)
…	…	…	…	…	…

After Table I was obtained, we are able to calculate the score by (1) and sort table rows in descending order by score value.

$$score = \sum_{i=1}^{n} \alpha_i p_i, \alpha_i \in [0,1], p_i = \{0,1,2\} \qquad (1)$$

For $n = 6$ and $6 \leq score \leq 8$, we have a "window" marked in Table I, by thick rectangle. By changing the values that constrain the score, we are able to slide the window up or down. We have three groups of students: the group above the window are students that are almost there and, usually they are able to figure it out by themselves, while the group of student below the window are students who need to study harder. So, desired position was to find a group that needs a "little push" to understand these concepts.

Except for the case when the student knows concepts to the maximum extent, by "window" we singled out five cases of cumulative understanding of these six concepts. These five cases are presented in form of "radar maps" that are easy to understand and read. The concepts are arranged in a clockwise direction, which corresponds to their order. Fig. 5 presents the case when students understand the concept of Pointer and memory allocation in maximal extend (high), while they are not able to understand how to create sub-nodes (low), while there is a lack of maximal understanding of all other concepts (medium).

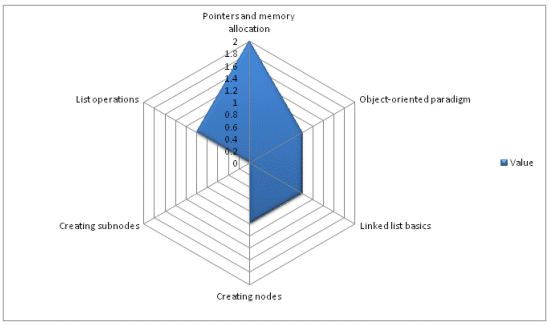


Figure 5: Pointer and memory allocation

Fig 6. represents students that are able to understand OO paradigm, as well as Creating nodes in the maximal extent, but understanding of the List operations is lacking (low).

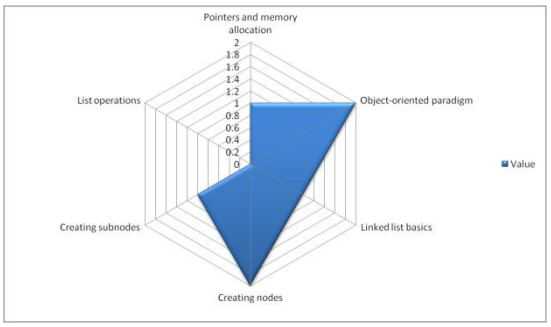


Figure 6: OO paradigm and creating nodes

Fig. 7 represents students who understand Pointers and memory allocation and Linked lists basics, however the understanding of other concepts is not maximal. There is no total lack of understanding of any concept.

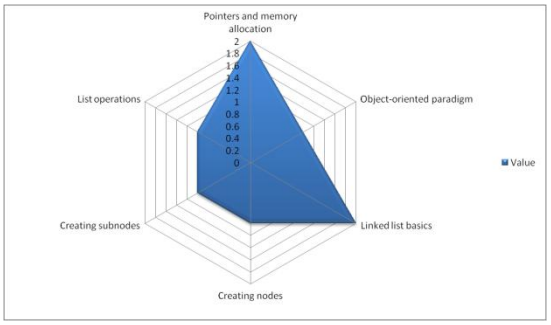


Figure 7:Pointers and memory allocation and List basics

Fig. 8 represents the group of students who understand three concepts in maximal extent, but they are not able to cope with OO paradigm.

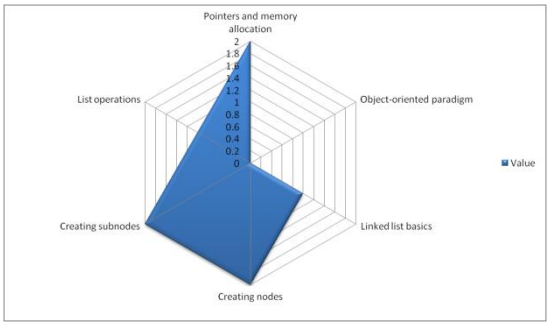


Figure 8: Pointers and memory allocation and Nodes

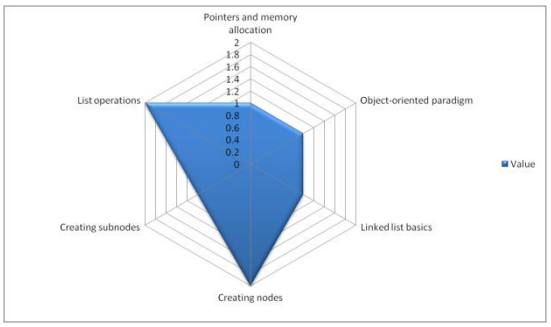


Figure 9: List operations and Creating nodes

Finally in Fig. 9 is represented a group of students who understand how to Create nodes and List operators, while other knowledge is lacking. There is no total lack of understanding of any concept.

IV. CONCLUSION

In this paper we show how object-oriented design can be applied to the implementation of a multi linked lists with a mixture of explanations, figures and sample codes. Linked lists are useful to study for two reasons. Most obviously, linked lists are a data structure which you may want to use in real programs. Somewhat less obviously, linked lists are great way to learn about pointers. Linked list problems are a nice combination of algorithms and pointer manipulation. Traditionally, linked lists have been the domain where beginning programmers get the practice to really understand pointers.

This paper is useful if you want to understand linked lists or if you want to see a realistic, applied example that uses structures inside classes and template classes.

We propose the exact way to estimate and visualize the extent of student's understanding of multi linked list in Object oriented C++ programming. This is a good starting point for further analysis of how students understand the basic concepts related to understanding of C++ linked lists. Six relevant basic concepts which are necessary for the understanding of C++ linked lists were defined by extensive literature review, while five group of students was formed in exact manner from empirical data. Six basic concepts are: Pointers and memory allocation, Object-oriented paradigm, Linked list basics, Creating nodes, Creating sub-nodes and List operations. The student's knowledge of basic concepts is rated as high (2), medium (1) or low (0).

First group of students is characterized by maximal understanding of Pointer and memory allocation, while they do not know how to create Sub-nodes of a C++ list. According to practical experience, they are able to recover through understanding of List operations. Second group of students is good in OO programming and Crating nodes, while they do not know how to implement List operations. These students are able to recover thanks to understanding of Node and Sub-node creation. Third group of students consists of students who understand each concept, although not with the maximal measure. Some backtracking to previous concepts is needed in order to fully understand C++ list implementation. Fourth group are students who lack in understanding of OO paradigm, but they are able to understand the implementation of C++ list, so their recover is possible by backtracking to OO paradigm. Finally, fifth group of students are those who understand each concept, but not with the maximal measure, so that backtracking to previous concepts is needed. According to our experience, the student who belongs to any of these five groups will be able to recover to maximal extent of C++ list understanding.

Some statistical analysis of presented approach is in progress so, future work will include more exact methods for student's knowledge assessment.

ACKNOWLEDGMENT

Ministry of Science and Technological Development, Republic of Serbia financially support this research, under

the project number TR32044 ”The development of software tools for business process analysis and improvement”.

REFERENCES

[1] Parlante, N. “Linked list basics”, Document #103, Standford CS Education Library, 2001.

[2] Stojisavljević, Đ. Brtka E. “Application of multi linked lists technique for the enhancement of traditional access to the data”, Proceedings of the International Conference on Applied Internet and Information Technologies, pp 403-407, Zrenjanin, Serbia, 2013.

[3] Dittrich, K. “Object-Oriented Systems – the notation and the issues”, International Workshop in Object-Oriented Database Systems, Pacific Grove, CA, 1986.

[4] Parlante, N. “Linked list problems”, Document #105, Standford CS Education Library, 2001.

[5] Tanenbaum A. Augenstein M. Langsam Y. “Data structures using C and C++”, PHI Learning, 2009.

[6] Della Penna G. “A type system for static and dynamic checking of C++ pointers”, Computer Languages, Systems & Structures 31, pp. 71–101, 2005.

[7] Woongsik C. and Kwang-Moo C. “Cycle elimination for invocation graph-based context-sensitive pointer analysis”, Information and Software Technology 53, pp. 818–833, 2011.

[8] Tüzün E., Tekinerdogan B., Kalender M. E. and Bilgen S. “Empirical evaluation of a decision support model for adopting software product line engineering”, Information and Software Technology 60, pp. 77–101, 2015.

Ontological Model of the Standardized Secondary School Curriculum in Informatics

Milinko Mandić*, Zora Konjović**, Mirjana Ivanović***
* Faculty of Education, University of Novi Sad, Sombor, Serbia
** Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
*** Faculty of Science, University of Novi Sad, Novi Sad, Serbia
milinmand@gmail.com, ftn_zora@uns.ac.rs, mira@dmi.uns.ac.rs

***Abstract*— The paper proposes ontological model of standardized secondary school curriculum in informatics. The model was created based on the ACM K12 CS curriculum proposal using competencies designed for the secondary level of education. The base class of ontological model is *Competence* with two direct subclasses (*Knowledge* and *Skills*). Skills are represented by classes corresponding to the categories of the cognitive process dimension of the revised Bloom's taxonomy. In addition to standardization of curriculum relying on ACM K12 CS curriculum model, a machine-readable representation that facilitates manipulation of the curriculum through applications intended for specific users (teachers, experts, administrative bodies, etc.) is proposed.**

I. INTRODUCTION

Ontologies can describe learning domains from different perspectives, allowing for a richer description and retrieval of learning contents [1]. Due to its ability to represent curriculum in a machine understandable manner, and the features of reuse and share, ontological approach has become widely used for representing some of the curriculum forms [1] [2] [3] [4] [5].

In [6] the authors state that the ontology offers an „objective base on which to build a curriculum recommendation". They use the ontology to represent computing curriculum and propose several potential uses of the ontology in curricula representation (for the purpose of distinguishing among computing programs and for highlighting the corresponding concepts in accordance with the selected outcome). In the paper [7] ontology has been created that allows the sharing of digital content for teaching mathematics and represents the mathematical domain of topics and skills. Reference [8] describes the development of a ontology-based curriculum knowledgebase which addresses the complexity of the interrelationships between the component parts of undergraduate enquiry based learning in medicine and other structured curricula and provides an approach to their collaborative maintenance. The same authors emphasize the importance of involving academics, students, teachers in the maintenance and improvement of a curriculum. Ref. [1] presents a system that has been developed with secondary school teachers that uses ontologies to support the development and management of the educative curriculum. In [3] the Bologna Ontology is created to model an academic environment as proposed by the Bologna reform. In [5], the model for representing ACM CS curricula based on the IEEE RCD standard is shown.

II. THE ONTOLOGICAL MODEL OF CURRICULUM

A. The class Competence

Informatics curricula have to be compliant at all the levels of education [5]. For example, higher education curricula for informatics teachers should be designed in the way that the graduated informatics teachers' *competencies* satisfy the needs of the current elementary and secondary school curricula [5]. After graduating from secondary school a student has to be educated enough in order to attend informatics courses of an adequate study program at higher education level. It is also important that secondary schools and faculties provide students with informatics *competencies* in order to satisfy companies' requirements for specific work posts.

Therefore, the ontological model of a secondary informatics curriculum is based on competencies, and the main class of the ontology is *Competence*.

There are numerous definitions of the term competence:

- A mixture of knowledge, skills, abilities, motivations, beliefs, values, and interests [9]
- A knowledge, skill, ability, or characteristic associated with high performance [10]
- It can be used to capture information about a skill, knowledge, ability, attitude, or learning outcome [11].

In [12], the authors list some examples of experts' efforts to define the term: "The knowledge, skills, and attributes that differentiate high performers from average performers", "It is a construct that helps define level of skill and knowledge", etc.

The same source concludes that "the term competence defines successful performance of a certain task or activity, or adequate knowledge of a certain domain of *knowledge* or *skill*".

Therefore, in this paper, the knowledge and skills mapped to specific classes of an ontological model of the curriculum (*Knowledge* and *Skills*), are represented as direct subclasses of *Competence*. (Figure 1). Classes *Skills* and *Knowledge* are related via object property *hasKnowledge*, that is its inverse property *hasSkill*.

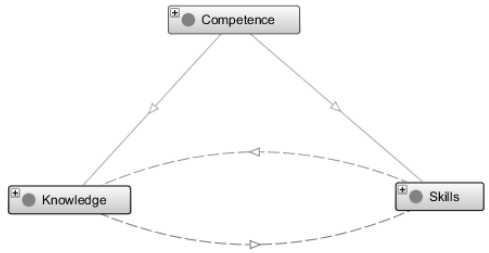


Figure 1. Structure of the class *Competence*

To ensure interoperability with learning management systems that provide information about competence, upper ontology classes are modelled in accordance with the IEEE RCD standard. The ontological representation of competencies based on the IEEE RCD standard (Figure 2) relies on the competences representation shown in [5] and [13].

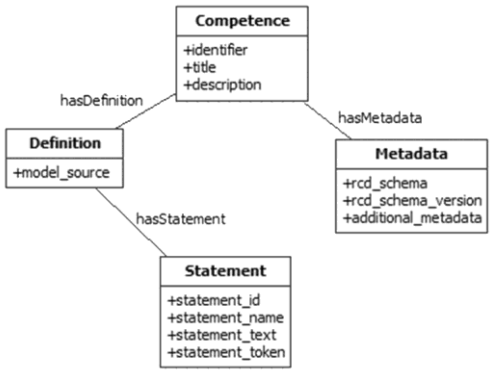


Figure 2. Ontological representation of the IEEE RCD standard [5]

The basic class is *Competence* with the properties *identifier, title and description*. Given that in the standard, these fields are defined as non-structured, they are represented as datatype properties in the ontological model. The field *definition*, although optional, is a structured part of the competence and can contain one or more *statements*, such as assessment criteria, outcomes, etc. It is modelled by the class *Definition* associated with the class *Statement* via the *hasStatement* object property. The *metadata* field is modelled by the *Metadata* class.

According to [11] the *Metadata* class enables the mapping of learning objects defined, for example, according to the IEEE LOM standard, while the classes *Definition* and *Statement* can be used for mapping data based on the principles for the assessment of student achievement, applied instructional methods, certification of competence, criteria, etc. Considering that the main goal of the ontological model curriculum presented in this paper is the defining of competencies from the perspective of acquired knowledge and skills, the classes *Metadata*, *Definition* and *Statement* are not structured in the ontology representing the secondary school curriculum.

Learning taxonomy is a way for describing the different behaviours in the learning process and the characteristics that students should develop throughout this process [14]. It provides a structure for the classification of educational goals and outcomes; in this paper it is used to define the class *Skills*. Skills are represented by classes corresponding to the categories of the cognitive process dimension of the revised Bloom's taxonomy [15], which is the dominant taxonomy in the area of Computer Science (CS) and in general [15] [16]. Exceptions are 'remember' and 'understand' categories, which are represented by a single class *Remember-understand,* due to the nature of having a CS domain in which the learning outcome involves only a recognition/memory without understanding is unlikely. Thus, the *Skills* subclasses are:

- *Remember-understand*,
- *Apply*,
- *Analyse*,
- *Evaluate* and
- *Create.*

B. Creating the ontology of the secondary school informatics curriculum

Secondary school informatics education clearly recognizes the need for and even provides the outcomes of curriculum standardization. The two most frequently cited models introducing curricula standardization in secondary school informatics education are the ACM K12 [17] and UNESCO/IFIP [18] models. ACM's K12 proposal is considered to be more modern and more comprehensive, because it proposes a curriculum based on computer science defined to represent a wider, more adequate and modern scientific discipline than ICT, as defined in [18].

For the above reasons, the ontological model of a secondary informatics curriculum in this paper was designed based on the ACM K12 CS curriculum proposal using only competencies designed for the secondary level of education (K8 or higher levels of standard).

Three general levels (L1, L2, L3) of the ACM K12 CS standard are separately described in detail in [19], [20] and [21] and consist of 12, 10 and 14 topics, respectively. Each topic contains a general description, a brief statement of support equipment for teaching, an assessment recommendations, detailed learning objectives, the focuses of each area and the proposal for the implementation of each of the focuses. Figure 3 shows part of the topic "Programming Languages" of the L2 level of ACM's proposal.

The ontological model of the secondary school informatics curriculum was created in two phases. Protégé tool [22] was used for the creation of the ontology. HermiT reasoner [23] was used for semantic verification of the model.

In the first phase, each of 36 topics of all three levels was modelled as a subclass of the *Knowledge* class. Focuses of the topics were modelled as the topic's subclasses. Learning objectives defined in [19], [20], [21] were mapped to the corresponding subclasses of the *Skills* class in accordance with the cognitive processes dimension of the revised Bloom's taxonomy.

For determining which general subclass of the *Skills* class certain objective belonged to, the synonyms of Bloom's taxonomy categories, as well as the detailed descriptions of the revised Bloom's taxonomy [24] [25] and digital taxonomy [26] were used. Learning objectives were related by the object property *hasKnowledge* with the appropriate topic that according to ACM references/documents they belonged to.

Other fields appearing in the curricula (teaching methods, knowledge assessment) are not currently covered by the ontological model. However, the model nonetheless enables mapping of these fields using the *Definition* and *Statement* subclasses. Additionally, learning objects can be easily incorporated into the model through the *Metadata* subclasses and connected via the object property with the appropriate thematic area.

Topic 11: Programming Languages

Topic Description:

Programming Languages will introduce the student to some basic issues associated with program design and development. The focus of this unit is to establish an appreciation of the work being done by software.

Textbooks and Supplies:

A programming language; interactive development environment recommended.

Time to Complete: 2-4 weeks

Student Learning Objectives	Assessment Measures
The student will be able to:	
1. Code, test, and execute a program that corresponds to a set of specifications.	Lab activity
2. Convert a word problem into code using top-down design.	Written activity Lab activity
3. Select appropriate data types.	Written activity Lab activity
4. Write structured program code.	Lab activity
5. Draw a series of diagrams showing the scope and values of variables during execution of a simple program.	Written activity

Assessment Recommendations: An average of 60% from combined assessment measures is required to demonstrate proficiency in course material.	
Lab activities	50%
Written activities, including tests, quizzes, and written assignments	50%

Detailed Outline	
Focus	**Sample Lab / Hands-on Activity**
1. Terminology	Identify and define key terms associated with programming.
2. Representation of text inside the computer	Each student writes a sentence in binary and exchanges it with a neighbor. The neighbor translates the sentence into text. Students stand or sit to mime a secret word in binary. Flashlights can also be used to represent binary code.
3. Representation of numbers inside the computer, including the largest and smallest values which can be represented in each of several types	Numbers are placed into imaginary bytes in a grid, each imaginary byte having a unique address. (A spreadsheet can be used for this purpose.) Instructions are provided to add and subtract values by address. Some of the resulting numbers should be too large to store in the imaginary byte and will overflow.

Figure 3. Topic 11: 'Programming Languages'

Thus, for example, the topic 'Programming Languages' (Figure 3) was mapped in the subclass of *Knowledge* class, while the topic's focuses were mapped in the subclasses of *Programming_Languages* class. The objective 'Write structured program code', in accordance with the revised Bloom's taxonomy, was modelled as a subclass of *Create* class and was associated with the class *Programming_languages*. The objective 'Select appropriate data types' was modelled by the subclass of *Analyze* class and was associated with the class *Programming_Languages*.

In the second phase, 36 topics of all three levels were placed into 13 areas that were determined based on overlapping of the focuses and goals in different topics. Topics and focuses with overlapping were labelled as 'related'. Then, related topics were mapped to one class and/or represented as subclasses of a common superclass. Characteristic examples are three topics: 'Problem solving', 'Algorithms', 'Problem solving and algorithms'. The parent classes *Problem_solving* and *Algorithms* became subclasses of the class *Problem_solving_and_algorithms* and the focuses belonging to the topics 'Problem solving' and 'Algorithms' were mapped to the corresponding subclasses of the class *Problem_solving_and_Algorithms*.

Related focuses of different topics were mapped to subclasses of a single superclass or to the superclass to which the related topic was mapped in the case where focus is represented by some special related topics (it appeared that some focuses of the lower-level ACM K12 curriculum (L1 or L2) were represented by special related topics in L2 or L3). An example is the focus 'Computing careers' of L1 level, which has been mapped to the *Careers_in_computing* superclass, as in L2 and L3 there is a special topic with the same name (Figure 4).

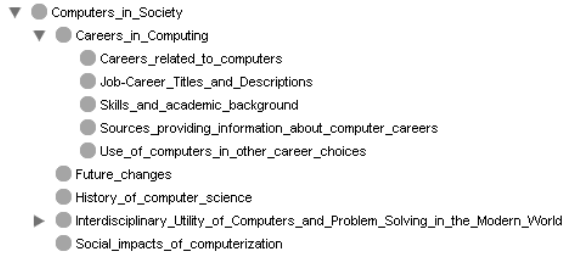


Figure 4. Structure of the class *Careers_in_computing*

Additionally, if some focuses have been repeated in several topics (as has, for example, arrays), they are mapped to a single class. Thus, the resulting final list of subclasses of the *Knowledge* class is shown in Figure 5.

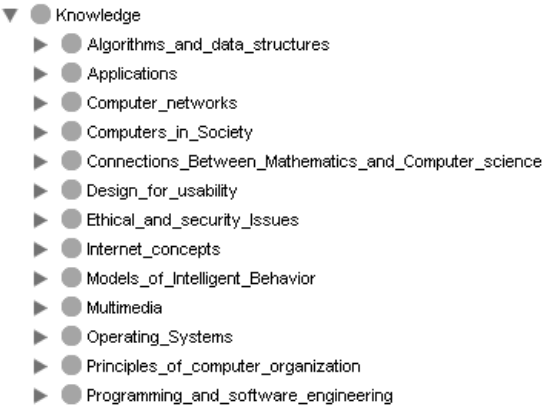


Figure 5. Subclasses of the class *Knowledge*

In order to include the latest topics in the curriculum, all the objectives defined in the latest current integrated version of the standard [17] were analysed and if some of them were not included in [19], [20] or [21], then the objective was added to the model and associated with the appropriate thematic area (subclass of *Knowledge* class). The reason is that in [17] learning objectives of students upon completion of specific levels of K12 curriculum are primarily defined, without sufficient explicit information about the required topics and knowledge that would enable consistent mapping into *Knowledg*e subclasses of ontological model.

Examples of the learning objectives from [17], added to the ontological model, are:

- Evaluate what kinds of problems can be solved using modeling and simulation,
- Evaluate algorithms by their efficiency, correctness, and clarity,
- Use mobile devices/emulators to design, develop, and implement mobile computing applications,
- Use Application Program Interfaces (APIs) and libraries to facilitate programming solutions,

- Demonstrate concurrency by separating processes into threads and dividing data into parallel streams.

Listing 1 presents a part of the owl code of the proposed ontological model.

```
  <owl:Class
rdf:about="#Code_a_program_that_corresponds_to_a
_set_of_specifications">
      <rdfs:subClassOf>
        <owl:Restriction>
          <owl:onProperty>
            <owl:ObjectProperty
rdf:about="#hasKnowledge"/>
          </owl:onProperty>
          <owl:someValuesFrom>
            <owl:Class
rdf:about="#Programming_Languages"/>
          </owl:someValuesFrom>
        </owl:Restriction>
      </rdfs:subClassOf>
      <rdfs:subClassOf rdf:resource="#Create"/>
    </owl:Class>
  <owl:Class
rdf:ID="Convert_between_decimal_binary_and_hexad
ecimal_numbers">
      <rdfs:subClassOf>
        <owl:Restriction>
          <owl:someValuesFrom>
            <owl:Class
rdf:about="#Connections_Between_Mathematics_and_
Computer_science"/>
          </owl:someValuesFrom>
          <owl:onProperty>
            <owl:ObjectProperty
rdf:about="#hasKnowledge"/>
          </owl:onProperty>
        </owl:Restriction>
      </rdfs:subClassOf>
      <rdfs:subClassOf rdf:resource="#Apply"/>
    </owl:Class>
```

Listing 1. A part of curriculum ontology

The ontological model is at the following address: www.pef.uns.ac.rs/SecondaryInformaticsCurriculum/index.html.

III. CONCLUSION

The main contribution of this paper is the ontological representation of a standardized secondary school informatics curriculum.

The advantages herein are the standardization of curriculum relying on ACM K12 CS, and machine-readable representation of the curriculum that facilitates curriculum manipulation through applications intended for specific user groups (teachers, experts in the domain field, the administrative bodies responsible for education management, etc.).

The simplicity of the model, which limits representation of the outcomes, representation of courses/topics prerequisites and aspects of the curriculum that are not closely related to the competencies, such as instructional method, methods of assessment, learning objects and the like, could be considered a model deficiency.

However, this deficiency may be mitigated by enriching the ontological model, which is one of the suggested future research directions.

For example, the model can be extended by alternatives for the outcomes' representations, which means adding new taxonomies in addition to the revised Bloom's taxonomy.

Introduction of the transitive object relation "prerequisite" would allow mapping information on the prerequisites for the study of a specific topic or course. In the existing curricula, this relationship is usually implicitly defined through the year/level of study, which makes it possible to establish preliminary links that can later be refined by manual intervention and/or machine learning methods applied to the courses/topics content.

REFERENCES

[1] .J.T. Fernández-Breis, D. Castellanos-Nieves, J. Hernández-Franco, C. Soler-Segovia, M. C. Robles-Redondo, R. González-Martínez, and M. P. Prendes-Espinosa, „A semantic platform for the management of the educative curriculum," *Expert Systems with Applications*, vol. 39, no. 5, 2012, pp. 6011-6019.

[2] Y.L. Chi, "Ontology-based curriculum content sequencing system with semantic rules," *Expert Systems with Applications*, vol. 36, no. 4, 2009, pp. 7838–7847.

[3] G. Demartini, I. Enchev, J. Gapany, and P. Cudré-Mauroux,"The Bowlogna Ontology: Fostering Open Curricula and Agile Knowledge Bases for Europe's Higher Education Landscape," *Semantic Web - Interoperability, Usability, Applicability*, vol. 4, no. 1, 2013, pp. 53-63.

[4] A. Elsayed, "Interaction with Content through the Curriculum Lifecycle," *Advanced Learning Technologies, ICALT 2009*, 2009, pp. 730 – 731.

[5] M. Mandić, M. Segedinac, G. Savić, and Z. Konjović,"IEEE RCD standard based ontological modeling of Computer Science curriculum," *Proceedings of the 3th International Conference on Information Society and Techology*, 2013, pp. 189-285.

[6] L. Cassel, G. Davies, R. LeBlank, L. Snyder, and H. Topi, "Using Computing Ontology as a Foundation for Curriculum Development," *Proc. SWEL@ITS '08: The Sixth International Workshop on Ontologies and Semantic Web for E-Learning*, 2008, pp. 21-30.

[7] P. Libbrecht, C. Desmoulins, C. Mercat, C. Laborde, M. Dietrich, and M. Hendriks, "Cross-curriculum search for intergeo.," In: S. Autexier, J. Campbell, J. Rubio, V. Sorge, M. Suzuki, & F. Wiedijk (Eds.), Intelligent computer mathematics, *Lecture Notes in Computer Science*, Vol. 5144/2008, 2008, pp. 520–535.

[8] H. Dexter and I. Davies, "An ontology-based curriculum knowledgebase for managing complexity and change," *Ninth IEEE International Conference on Advanced Learning Technologies*, *ICALT*, 2009, pp.136-140.

[9] E. A. Fleishman, L. I. Wetrogan, C. E. Uhlman, and J. C. Marshall-Mies, „Abilities," In N. G. Peterson, M. D.,Mumford, W. C. Borman, P. R. Jeanneret, and E. A. Fleishman (Eds.), *Development of prototype occupational information network content model* , Salt Lake City, UT: Utah Department of Employment Security, vol. 1, 1995, p. 1086.

[10] R. Mirabile, "Everything you wanted to know about competency modeling," *Training and Development*, vol. 51, no. 8, 1997, pp. 73–77.

[11] *IEEE Standard for Learning Technology—Data Model for Reusable Competency Definitions*, Learning Technology Standards Committee of the IEEE Computer Society, 2008. http://www.doleta.gov/usworkforce/pdf/2007-ieeecomp.pdf.

[12] J. Shippmann, R. Ash, M. Battista, L. Carr, L. Eyde, B. Hesketh, J. Kehoe, K. Pearlman, E. Prien, and J. Sanchez, "The practice of competency modeling," *Personnel Psychology*, vol. 53, no. 3, 2000, pp. 703-740.

[13] J. De Coi, E. Herder, A. Koesling, C. Lofi, D. Olmedilla, O. Papatreou, and W. Siberski, "A Model for Competence Gap Analysis," *Proceedings of 3rd International Conference on Web Information Systems and Technologies (WEBIST)*, 2007.

[14] G. O'Neill and F. Murphy, "Guide to Taxonomies of Learning," *UCD Teaching and Learning/ Resources*, 2010. http://www.ucd.ie/t4cms/ucdtla0034.pdf

[15] U. Fuller, et al., "Developing a Computer Science-Specific Learning Taxonomy," *ACM SIGCSE Bulletin*, vol. 39, no. 4, pp. 152-170, 2007.

[16] L.W. Anderson, D.R. Krathwohl, P.W. Airasian, K.A. Cruikshank, R.E. Mayer, P.R. Pintrich, J. Raths, and M.C. Wittrock, *A taxonomy for Learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives,* New York: Addison Wesley Longman, 2001.

[17] CSTA K–12 Computer Science Standards, The CSTA Standards Task Force, 2011. http://csta.acm.org/Curriculum/sub/CurrFiles/CSTA_K-12_CSS.pdf

[18] *INFORMATION AND COMMUNICATION TECHNOLOGY IN SECONDARY EDUCATION, A Curriculum for Schools,* UNESCO, 2000. http://wwwedu.ge.ch/cptic/prospective/projets/unesco/en/curriculum2000.pdf

[19] A. Verno, D. Carter, R. Cutler, M. Hutton, and L. Pitt, "A Model Curriculum for K-12 Computer Science: Level II Objectives and Outlines," 2004. http://csta.acm.org/Curriculum/sub/CurrFiles/L2-Objectives-and-Outlines.pdf

[20] B. Madden, A. Verno, D. Carter, S. Cooper, T. Cortina, R. Cudworth , B. Ericson, and E. Parys, "A Model Curriculum for K-12 Computer Science: Level III Objectives and Outlines," 2007. http://csta.acm.org/Curriculum/sub/CurrFiles/L3-Objectives-and-Outlines.pdf

[21] D. Frost, A. Verno, D. Buckhart, M. Hutton, and K. North, "A model curriculum for K-12 Computer Science: Level I Objectives and Outlines," 2009. http://csta.acm.org/Curriculum/sub/CurrFiles/L1-Objectives-and-Outlines.pdf.

[22] N. F. Noy, R. W. Fergerson, and M.A. Musen, "The knowledge model of Protege-2000: Combining interoperability and flexibility," *Proceedings of the 12th International Conference on Knowledge Engineering and Knowledge Management (EKAW'00)*, 2000, pp. 17-32.

[23] R. Shearer, B. Motik, and I. Horrocks, "HermiT: A Highly -Efficient OWL Reasoner," *Proceedings of the OWL Experiences and Directions Workshop*, 2008.

[24] D. R. Krathwohl, „A revision of bloom's taxonomy: An overview," *Theory into Practice*, vol. 41, no. 4, 2002, pp. 212-218.

[25] R.Heer, "A Model of Learning Objectives–based on A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives," Center for Excellence in Learning and Teaching, Iowa State University, 2012. http://www.celt.iastate.edu/teaching-resources/effective-practice/revised-blooms-taxonomy/

[26] A. Churches, "Bloom's Digital Taxonomy," 2007. http://www.techlearning.com/techlearning/archives/2008/04/andrewchurches.pdf

Architecture and Algorithms for Filtering Tweets Based on Chosen Countries and Cities

Nemanja Igić*, Vladimir Dimitrieski*, Ivan Luković *, Slavica Aleksić*, and Milan Čeliković*
* University of Novi Sad/Faculty of Technical Sciences, 21000 Novi Sad, Serbia
{nemanjaigic, dimitrieski, ivan, slavica, milancel}@uns.ac.rs

***Abstract*— In this paper we present an algorithm for filtering Twitter data based on the tweet geographic location. Desired geographic locations are provided as a set of parameters, and different properties of a tweet are considered to determine the location. A user may also choose the number of threads and amount of memory used in filtering process. In this way, the user may fine-tune the algorithm performance. Filtered data are stored in the Hadoop distributed file system which runs on a 16 nodes cluster. Therefore, the analysis may be executed in a distributed environment. We also present system architecture which supports the filtering process and analysis of filtered data.**

I. INTRODUCTION

A social network is a structure composed of a series of social actors and a set of connections between these actors. Social actors can be individuals or organizations. Among the most popular social networks today are Google+ with 1.6 billion members, Facebook with 1.28 billion members, Twitter with 645.75 million members and Qzone with 480 million members [1]. Social networks produce large amount of data from different geographic areas on a daily basis making them a perfect platform for such analysis. In this paper, we observe data gathered from Twitter. Twitter is an online social networking service that enables users to send and read short 140-character messages called "tweets" [2]. Twitter provides a rich REpresentational State Transfer (REST) Application Programming Interface (API) allowing programmatic access to read and write Twitter data [3].

The main problems in analyzing data from social networks are that data cover a wide range of topics and there is a lack of metadata information relevant to the analysis (geographic coordinates, tags, etc.). Also, metadata is mostly user-defined, and in many cases caries inaccurate information. To provide accurate analysis, there is a need for collecting large amounts of data and providing a way to compensate for missing metadata. The main motivation behind this work is a lack of efficient algorithms for finding tweet locations when there are no geographic coordinates specified, which are specified in less than 2 percent of tweets [4]. Our goal is to implement a fast algorithm for filtering tweets based on predefined country and city parameters with acceptable accuracy. This kind of filtering may be useful for various analysis, since one will gather approximately 40 times more tweets from the specified locations rather than using only twitter geo tags, as it will be presented further in this paper.

As the amount of collected data grows, there is a need for an easy way to provide horizontal scalability [5]. We choose Hadoop for this purpose. Hadoop is a software framework for distributed storage and distributed data processing on commodity hardware clusters. It represents an implementation of the Google system for executing MapReduce (MR) programs presented in [6]. Hadoop stores data in Hadoop Distributed File System (HDFS) [7]. HDFS represents a distributed file system that has a high resistance to failures and is designed to be used on commodity hardware. This framework is based on the MapReduce programming model [8], which aims at processing large amounts of data through parallel and distributed algorithms running on a cluster. We chose Hadoop since it is an open source system which distributes tasks on all nodes within the Hadoop cluster. Therefore, as amounts of data grow, the performance would not be degraded as extra nodes can be always added to the cluster to make the analysis faster.

In this paper we present an algorithm for filtering data gathered from Twitter. The algorithm provides two processes. One, that fetches tweets provided via Twitter API and second, filtering process that filters data by defined cities and countries. The filtering process is executed in parallel. Filtered data are stored in HDFS for further analysis. In the paper we also present architecture of the cluster where the filtered data are stored. Also, we discuss the influence of a system architecture design on the system performance. The system architecture is based on a cluster with 16 nodes, each with a Hadoop instance installed.

In addition to Introduction and Conclusion, the paper is organized in five sections. Related Work is presented in Section II. In Section III, we present the proposed architecture. In Section IV we describe a part of a tweet structure relevant for the proposed algorithm. The algorithm used for filtering data collected from Twitter by the geographic location is presented in Section V. In the same section we discuss the storing mechanism for data in HDFS. In Section VI we present an example of filtering data using the algorithm presented in Section V.

II. RELATED WORK

To the best of our knowledge, there is only one paper that presents algorithm for filtering tweets by location. Rest of the papers mentioned in this section present similar algorithms and architectures to this paper. However, none of these papers provides parallel, real-time location filter.

Mahmud et al. [9] use an ensemble of statistical and heuristic classifiers to predict tweet locations based on the content of users‟ tweets and their tweeting behavior. Chen et al. [10] analyze user tweets through brand-specific intelligent filters to group users based on specific opinion. Kapanipathi et al. [11] used a Semantic Web

approach to filter public tweets matching interests from personalized user profiles. Mane et al. [12] used to analyze tweets using Hadoop, but without filtering and grouping them before analysis. Siddaraju et al. [13] present different approaches of processing big data with Hadoop in general. In contrast to all these works, we use a parallel algorithm to find tweet locations in real-time without a need to train and use classifiers. Only filtered data are stored in HDFS. Therefore we keep only tweets that are relevant to our analysis.

III. SYSTEM ARCHITECTURE

In order to achieve good Hadoop performances, a system based on Hadoop must be set up in distributed environment. For the purposes of this research, we have used an infrastructure consisting of 16 networked computers, as a basis of our system architecture.

The architecture of our system comprises a cluster consisting of 16 nodes. On all nodes we have installed Fedora 19 [14] operating system. Each node has 230 GB of dedicated memory for Fedora 19, of which 180 GB is reserved for Hadoop. On each node there is 4 GB RAM available and four processing cores. The nodes are a part of a single computer network with bandwidth of 100 Mb/s. Within the network, all computers are uniquely identified by their hostname formed by concatenating the letter „S" with a number from the interval [201, 216]. On all computers, Hadoop version 2.4.1 is installed. This version is chosen as it is the latest stable version at the time of writing. Hadoop instances are configured following tutorials presented in [15] and [16]. We chose computer with hostname S206 to be our master node purely for physical accessibility reasons. Master node is responsible for distributing tasks to slave nodes within the cluster where Hadoop is installed. Hadoop instances share data via Secure Shell (SSH) protocol [17]. Architecture organized in this way facilitates horizontal scaling. Therefore, as amounts of collected data grow, by adding extra nodes to our cluster, performance will not degrade. In Figure 1, we present our system architecture.

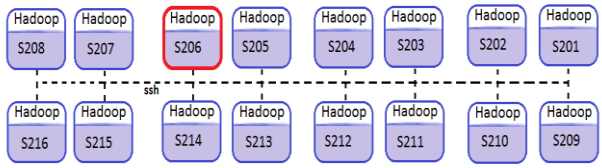


Figure 1The system architecture

IV. TWEET STRUCTURE

In this section we present a part of the tweet structure that is used for the filtering algorithm that will be presented in the next section. The entire specification of the tweet structure may be found in [18].

Data fetched from the Twitter data stream may be categorized in one of the following categories: tweets, messages about erased statuses and notifications about exceeded number of connection attempts. From all of these categories, in this paper we only consider tweets. Data from other categories are ignored because they do not carry information relevant to the analysis performed in this paper.

Fields from the tweet structure which have been used in our research are:

- *coordinates* - represent the geographic location of a tweet as reported by the user or client application.
- *lang* - when present, indicates a language identifier corresponding to the machine-detected language of the Tweet text, or "und" if no language could be detected.
- *user* - the user who posted this tweet. A location and a time zone are defined by a user in the *location* and *time_zone* fields.
- *text* - The actual UTF-8 text of the tweet.

Example of the part of the tweet structure described above with values of aforementioned fields is presented in Listing 1.

```
"coordinates":
{
    "coordinates":
    [
        -75.14310264,
        40.05701649
    ],
    "type":"Point"
}
"lang": "en"

"place":
{
    "attributes":{},
     "bounding_box":
    {
        "coordinates":
        [[
                [-77.119759,38.791645],
                [-76.909393,38.791645],
                [-76.909393,38.995548],
                [-77.119759,38.995548]
        ]],
        "type":"Polygon"
    }
}
"text": "Tweet Button, Follow Button, and Web
   Intents javascript now support SSL
   http:\/\/t.co\/9fbA0oYy ^TS"
"user":
{
    "location":"San Francisco, CA",
    "time_zone":"Pacific Time (US &amp;
      Canada)"
}
```

Listing 1 Example of the tweet structure part used in our approach

V. COLLECTING, FILTERING AND STORING TWEETS

In this section, we present the algorithm for collecting, filtering and storing data, which represents the central part of our system. First, we give a brief overview of the system architecture component which main task is to collect, filter and data from twitter and store it in HDFS. Second, we present the algorithm for filtering data based on the tweet structure presented in Section IV. Finally, we present storing mechanism of filtered data in HDFS.

A. A Component for Collecting, Filtering, and Storing Data

Collecting, filtering and loading data is performed in parallel. The component for collecting, filtering and storing data (CFS) is implemented in Python

programming language. We chose Python as it has many libraries which facilitate efficient implementation of the component. For example, we used library for concurrent programming to implement concurrent execution of CFS component. JSON library was, in our case, used for transforming tweets fetched from the Twitter stream represented in JSON format to objects. In Figure 2 we present a structure of CFS that consists of three segments. In one thread, data are downloaded from the Twitter stream and placed in a queue for further processing. This is presented in Figure 2 in the top segment labeled "part 1". The queue used for this task is implemented in the Python Queue package, as thread safe collection. Other threads, responsible for filtering and storing data, are presented in Figure 2 in the middle segment, labeled "part 2". The number of threads responsible for filtering and storing data is defined by a parameter. Data are stored in a temporary list. The list size is also defined as a parameter. Once the maximum size of the list is reached, a thread responsible for storing data empties the list and transfers data from the list to HDFS. This is presented in Figure 2 in the bottom segment labeled "part 3".

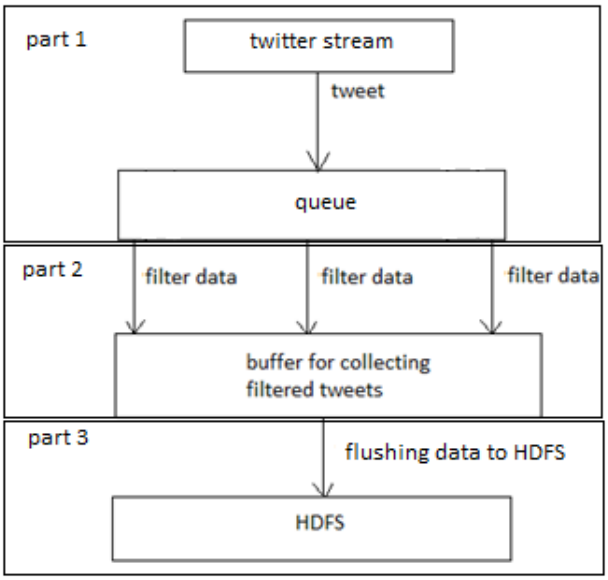


Figure 2 The CFS component

B. *Filtering Algorithm*

The main goal of the filtering algorithm is to provide a suitable way to extract tweets that are only from countries and cities relevant to the analysis. Countries and cities are defined as parameters in a parameter file. The format of the parameter file is presented in Section VI.

After a tweet is fetched from the Twitter stream, it is necessary to check whether the tweet originates from one of the defined countries. If a tweet is from one of the defined countries and cities it should be stored, otherwise it should be ignored. Data filtering is done in five steps presented in Figures 3 to 7. These steps are sorted by the relevance, where the first step provides the most accurate results, while the last step provides the least accurate results. Each successive step determines the country and city of a tweet with less accuracy. A next step in the algorithm is executed only after the tweet location cannot be determined by the previous step. This way of handling data may lead to inconsistency as the algorithm may not be able to precisely determine a location due to the fact that too much metadata is missing. Complete consistency in this case will result in a rejection of too much information, which would be later reflected on the reliability of the analysis. In the following paragraphs, each step will be explained in detail.

In the first step of our algorithm, a tweet is checked for the *coordinates* field. If the tweet contains the coordinate value, the next action is to check if its value corresponds to coordinates of one of the defined countries. If the coordinates belong to one of the defined countries, the distance between the given coordinates and the coordinates of each city from that country is calculated. A city that is closest to the given coordinates is declared to be the city from where the tweet is originated. Afterwards, country and city names are placed in the *location* field of the tweet. Modified tweet is then placed in a temporary list of collected tweets. In Figure 3 we present a part of the activity diagram for this step. If the aforementioned conditions do not hold, the algorithm proceeds to the second step.

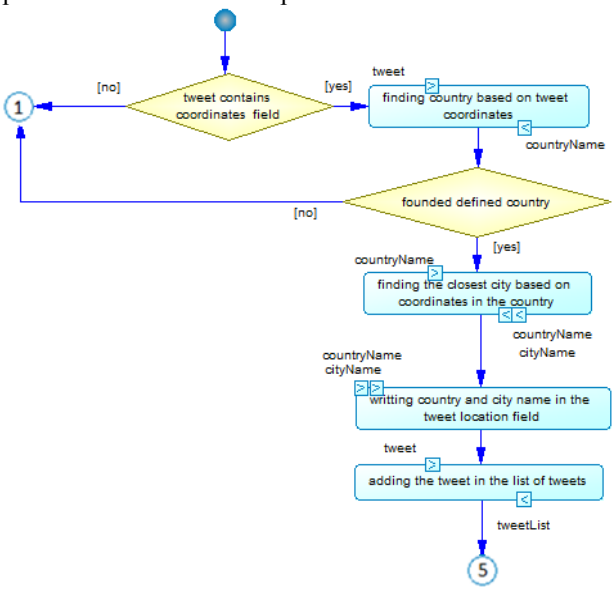


Figure 3 The first step of the filtering algorithm

In the second step of our algorithm, a tweet is checked for the *place* field. If the tweet contains the value of the *place* field, within the *place* field there is the *bounding_box* subfield that represents a polygon consisting of coordinates enclosing the place. The center of the bounding box is declared to be the geo-location of the tweet. The rest of the activities in this step are carried out analogously to the first step of the algorithm. In Figure 4 we present a part of the activity diagram for this step. If the aforementioned conditions do not hold, the third step of the algorithm is performed.

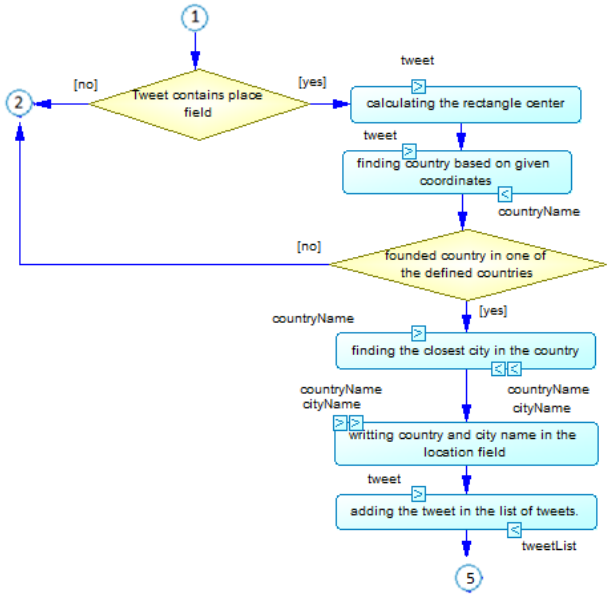


Figure 4 The second step of the filtering algorithm

In the third step of our algorithm, data are filtered by the *location* filed. First, words in the *location* field are parsed. This can be achieved by replacing all characters that are not alphanumeric with whitespace and then by splitting the resulting string by whitespace. Next, it must be checked whether some words include a variation of the name of defined city or country. Name variations are associated with defined country and city names and are defined manually as a parameter. They represent alternative or slang names for those countries and cities. Afterwards, a check for the city name variation is performed. If there is a match, country name from where the city is from and city name is placed in the *location* field of the tweet. In a case when there are no matches found for city name variations and there is a country name variation, only the country name is written into the tweet *location* field. In Figure 5 we present a part of the activity diagram for this step. If the aforementioned conditions do not hold, algorithm proceeds to the fourth step.

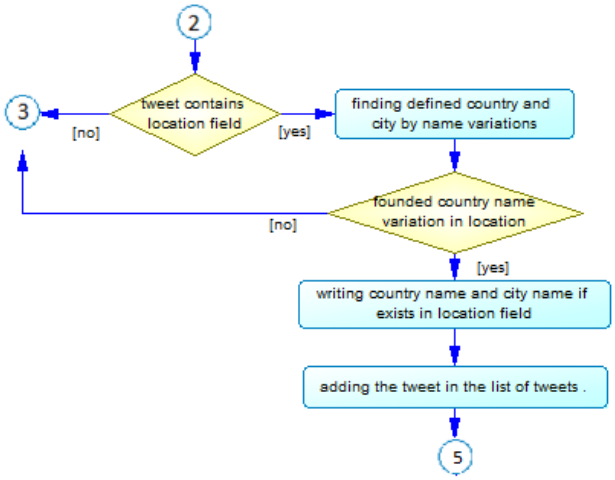


Figure 5 The third step of the filtering algorithm

The fourth step checks whether the language defined in the *lang* field, is a vernacular for one of the countries. Due to errors in the identification of the language in the *lang* field, mutually intelligible languages should be also considered as vernaculars. If the language in the tweet is one of the vernaculars of the defined country, it is assumed that a tweet is from the observed country and country name is written in the *location* field. In Figure 6 we present a part of the activity diagram for this step. If the previous conditions do not hold, algorithm proceeds to the fifth step.

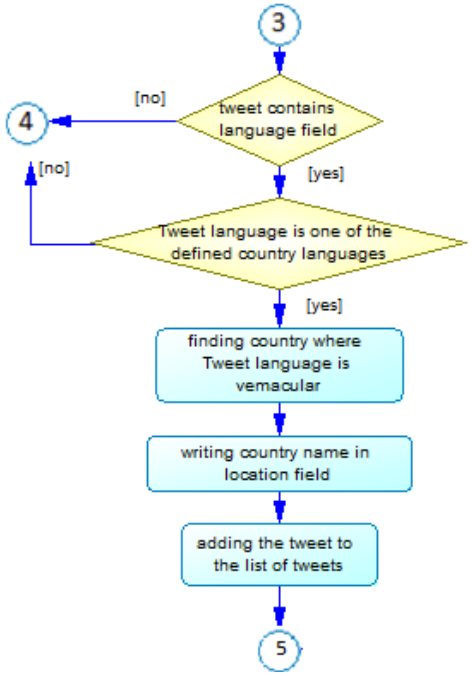


Figure 6 The fourth step of the filtering algorithm

In the final step of the algorithm, the value of the *time_zone* field is checked. The value of this field is parsed as in the third step. Then the country name variations are checked if they appear in one of the parsed words. If there is a country for which this condition is satisfied, the name of that country is written in the *location* field. If the previous conditions do not hold, algorithm finishes filtering given tweet. In that case, a thread executing the algorithm fetches a new tweet and repeats the process again. In Figure 7 we present a part of the activity diagram for this step.

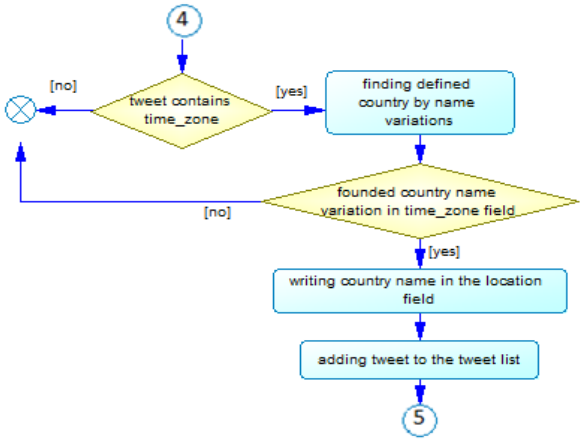


Figure 7 The fifth step of the filtering algorithm

Aforementioned steps are chosen because they reference fields from Twitter structure which contain information about tweet location. Steps are sorted by relevance by using results from analysis presented in

Section 6, where accuracy for each step was measured and by using explanation of results for each step which is also presented in Section 6.

C. Storing data in HDFS

After the list of collected tweets reaches the maximum size, the list values are stored in HDFS. Since storing data to the hard disk is expensive operation, the size of the list should be as large as possible (more than a hundred of elements).

Since Hadoop is based on the MapReduce programming model, data are stored as key-value pairs. After the list reaches defined size, it gets stored in HDFS. In our approach, the key in the key-value pair represents the *id* of the list, which values are incremented by 1, starting from 1. The value in the key-value pair represents the entire content of the list. After the list is stored in HDFS, list elements are removed. In Figure 8 we present a part of the activity diagram for storing data in HDFS.

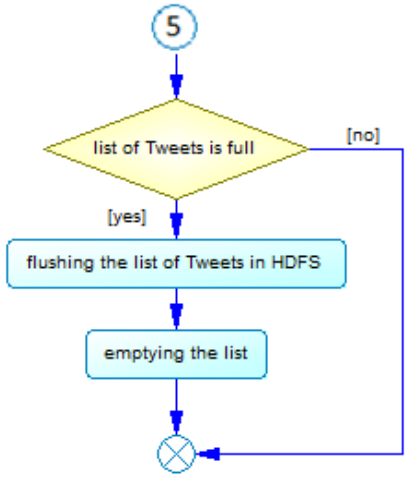


Figure 8 The algorithm steps for storing data in HDFS

Since the first step checks the *coordinates* field, one might be sure that tweet is originated from desired location. Therefore, values of fields used in step 3, 4 and 5, are checked and if certain values for those fields appear predefined number of times, they are also considered in the algorithm. This represents a way to handle synonyms (for example, in *location* field Novi Sad might often appear as Serbian Athens).

VI. Example of the System Usage

In this section we present an example of the system usage by collecting tweets which are originated from Serbia, Croatia, Montenegro or Bosnia and Herzegovina. First, we present set-up of necessary input parameters for this example. Second, we present the results of running the filtering algorithm for the defined set of parameters and the accuracy of the algorithm for this example.

In Listing 2, we present performance parameters. The RAM parameter defines the percentage of free memory that is to be allocated for the temporary list of tweets that gathers filtered tweets. The thread parameter defines the number of threads which are used for filtering data.

```
RAM#60
thread#5
```

Listing 2 Performance parameters

In Listing 3, we present parameters relevant to the cities we observe. First, we define country from where the city is from. Afterwards, we define a city name followed by geographic coordinates in a form of a pair (longitude, latitude). Finally, we define city name variations.

```
RS#Novi Sad#45.253829,19.830435#novi sad,нови сад
RS#Beograd#44.808182,20.452938#београд,beograd,belgrade
RS#Nis#43.325485,21.904346#ниш,nis,niš
ME#Podgorica#42.458755,19.253537#podgorica
BA#Sarajevo#43.861104,18.412421#sarajevo
HR#Zagreb#45.807778,15.976691#zagreb
BA#Banja Luka#44.778541,17.204558#banja luka
```

Listing 3 Parameters relevant to cities

In Listing 4, we present parameters relevant to the observed countries. The first parameter represents country code. Afterwards, we define city variation names, followed by country codes of vernaculars. In the end we define time zone name variations, which are usually comprised of name variations of the given country and cities. Those countries and cities represent time zone boundaries of countries and their subdivisions.

```
RS#србија,pc,срб,srb,srbija,serbia#rs,sr,sl#s
rbija,serbia,beograd,belgrade
BA#bosna,bosna i hercegovina,bih,republika
srpska,република српска#sl,ba#bosna i
hercegovina,bosnia and
herzegovina,sarajevo,sarajevo#
HR#hrvatska,hr,cro,croatia#sl,hr#hrvatska,cro
atia,zagreb
ME#crna gora,montenegro,mne,cg#sl,me#crna
gora,montenegro,podgorica
```

Listing 4 Parameters relevant to countries

Tweets were collected for 6 hours. In that time, the process defined by the algorithm presented in this paper filtered 2,213,098 tweets. 2,180 tweets were recognized as tweets from defined locations, which represented 0.0985% from the total number of tweets.

Our analysis was performed in two phases. First, we take a sample of 4,458 tweets from all gathered tweets. Our algorithm recognizes 5 tweets from that sample to be from defined territories. Afterwards, we manually check all 4,458 sample tweets. In this case there are 4 tweets that match defined countries and cities. One tweet from Campo Grande, city in Brazil, is recognized as Montenegrin due to the *location* field. The *location* field contains a value “cg” which represents acronym for this country name in their native language. However, the acronym of the Campo Grande city is also “cg”. Value “cg” represents country variation name in our case. Thus, we recognize all tweets for defined countries and cities in our sample, but one false positive. There are no false negatives in this example.

Second, we analyze how many tweets each step of our algorithm finds and what is the accuracy of each step in the aforementioned 2,180 tweets. The analysis is done manually, by looking the content of all the aforementioned tweets. In the first step, 58 tweets are found, which makes 2.6% of total number of tweets, and

had accuracy of 100%. This percentage of accuracy is to be expected, because geographic coordinates are specified. The only problem is that we group all tweets by cities which we define and therefore we declare that tweet is from the closest defined city. In our case it is desired behavior, because we want to group tweets in areas around largest cities of each country. In the second step, 29 tweets are found, which makes 1.3% of total number of tweets, with accuracy of 96%. Errors can appear to the following two reasons. First, the *place* field contains bounding box of place which is mentioned in tweet, not an exact geographic coordinate. Second, the *place* field contains places mentioned in tweet, not the actual place from where tweet is from. In the third step, 697 are found which makes 32% of total number of tweets. This makes with the accuracy of 94%. 80% of mistakes are made due to the value "cg" in the *location* field. The value "cg" represents Campo Grande, so to increase the accuracy of this step it is necessary to exclude value "cg" from parameters. Therefore, choice of defined parameters is very important for performance of this step, especially city and country variation names. In the fourth step 1,078 tweets are found with accuracy of 84%. Most of the errors are made due to language identification algorithm. Also, there is a problem with languages which are native languages in more than one country. This step varies for different countries and languages but must be included because in this case there are 49.4% of tweets found due to this step. In the final step, 318 tweets are found, which makes 14.6% of total number of tweets, with accuracy of 81%. The problem with this step is that users define their time zone manually and in many cases it is incorrect. Accuracy of the algorithm is calculated by Formula 1:

$$a = a1 * f1 + a2 * f2 + a3 * f3 + a4 * f4 + a5 * f5$$

Formula 1 Accuracy of the algorithm

where a is the accuracy of the algorithm, ai, $i \in \{1..5\}$, accuracy of step 1 to 5, and fi, $i \in \{1..5\}$, number of tweets for that step divided by total number of tweets. Accuracy for this example is 86%.

VII. CONCLUSIONS

In the paper we present an algorithm for filtering data gathered from Twitter and architecture for the system to analyze that data.

Data filtered by the algorithm may be used for various analyzes that are based on specific locations. Our algorithm showed good performance for the example presented in Section VI. However, the performance may vary for different values of input parameters. Some steps may show very low performance for some values of parameters, like it is the case with step four for the English language. Therefore, there is a need for a preprocessing algorithm, to find the most efficient combination of steps and for efficient way to provide performance analysis.

This platform can be a basis for a powerful decision support system in a form of a platform as a service. In our future research, we plan to add data from multiple new sources such as news web sites, forums, and blogs to collect more data from wider range of people and therefore to improve our analyses. We also plan to implement data mining packages for complex analysis of public opinion. Based on those data mining packages, we plan to build a Domain-Specific Language, which would allow users who do not have any experience in programing to specify and generate MapReduce programs to be executed on our system. With those features, we can provide a fast and accurate system for analyzing public opinion on various topics which can be used by users with no or little programing experience in a simple way.

ACKNOWLEDGMENT

The research presented in this paper was partially supported by Ministry of Education, Science and Technological Development of Republic of Serbia, Grant III-44010.

REFERENCES

[1] "Social network users" [Online] Available: http://en.wikipedia.org/wiki/List_of_social_networking_websites [Accessed: 29.11.2014].

[2] "Tweet definition" [Online], Available: http://whatis.techtarget.com/definition/Twitter [Accessed: 18.11.2014].

[3] "Twitter API" [Online], Available: https://www.dev.twitter.com/rest/public [Accessed: 18.11.2014].

[4] "Twitter geolocation and its limitations" [Online], Available: http://dfreelon.org/2013/05/12/twitter-geolocation-and-its-limitations/ [Accessed: 05.01.2015].

[5] C. Weinstock and J. Goodenough, "On System Scalability", in Software Engineering Institute (CMU/SEI-2006-TN-012), March 2006

[6] Jeffrey Dean and Sanjay Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters" in *Communications of the ACM Volume 51 Issue 1, January 2008*

[7] Ralf Lammel, "Google"s MapReduce Programming Model " in *Science of Computer Programming Volume 70 Issue 1,* January, 2008

[8] "Hadoop HDFS," [Online], Available: http://hadoop.apache.org/docs/r1.2.1/hdfs_design.html [Accessed: 20.11.2014].

[9] Jalal Mahmud, Jeffrey Nichols, Clemens Drews, "Home Location Identification of Twitter Users", in *ACM Transactions on Intelligent Systems and Technology (TIST) Volume 5 Issue 3,* September 2014

[10] Jilin Chen, Allen Cypher, Clemens Drews , Jeffrey Nichols, "CrowdE: Filtering Tweets for Direct Customer Engagements" in *Cameron Wynn,* September, 2014

[11] Kapanipathi, Orlandi, Sheth, Passant, "Personalized Filtering of the Twitter Stream" in *Marco de Gemmis*, November, 2011

[12] Sunil B. Mane, Yashwant Sawant, Saif Kazi, Vaibhav Shinde , "Real Time Sentiment Analysis of Twitter Data Using Hadoop " in *International Journal of Computer Science and Information Technologies, Vol. 5* March , 2014

[13] Siddaraju,, Sowmya , Rashmi, Rahul, "Efficient Analysis of Big Data Using Map Reduce Framework " in *International Journal of Recent Development in Engineering and Technology Volume 2, Issue 6*, June 2014

[14] "Fedora project" [Online] Available: http://www.server-world.info/en/note?os=Fedora_19&p=download [Accessed: 28.11.2014].

[15] Mihir Bellare, Tadayoshi Kohno, Chanathip Namprempre, "Authenticated Encryption in SSH: Provably Fixing the SSH Binary Packet Protocol", in *Ninth ACM Conference on Computer and Communications Security*, ACM, 2002

[16] "Single node cluster tutorial", [Online], Available: http://www.michael-noll.com/tutorials/running-hadoop-on-ubuntu-linux-single-node-cluster/ [Accessed : 21.11.2014].

[17] "Multi node cluster tutorial", [Online], Available: http://tecadmin.net/set-up-hadoop-multi-node-cluster-on-centos-6/ [Accessed : 21.11.2014].

Automatic data extraction from radargrams

Aleksandar Ristic, Aleksandra Radulovic, Miro Govedarica, Milan Vrtunski
University of Novi Sad, Faculty of Technical Sciences, Department of automation, geomatics and control systems
Trg Dositeja Obradovica 6, 21000, Novi Sad, Serbia

Abstract— **Radargrams are result of acquisition with GPR scanning technology. Most of underground utilities has cylindrical shape and is represented in radargrams by hyperbolic signatures. In this paper a new automated procedure for hyperbolic signatures detection and data extraction in radargrams is proposed. Extracted data is the set of points which represents vectorized form of hyperbolic reflections. Resulting data from automated procedure were later used to simultaneously estimate the radius of a cylindrical object and the wave propagation velocity based on fitted hyperbola geometries. Estimation of geometry parameters and wave propagation velocity from extracted set of points is a complete solution for underground utility and soil characterization. A number of experiments showed that the procedure is robust to various types of influences.**

I. INTRODUCTION

In recent period, GPR technology has become more accessible and more present in engineering applications. This led to an increased amount of data being collected with GPR [3]. In addition, interpretation of results from the radargram (e.g., underground utility detection) is a complex task in terms of operators' knowledge and skills. Therefore, the increased amount of data and complex interpretation are pre-conditions for the development of automated procedures for detection and interpretation of anomalies in radargrams.

From a technical point of view, software detection of anomalies (hyperbolic reflections, for instance) in radargram is difficult because:

- Various types of media surrounding the objects produces excessive data (caused by different geological environments and change of volumetric moisture content)
- Incomplete or noisy hyperbolic reflections (caused by conditions of acquisition and/or media inhomogeneities)
- Interference of neighboring hyperbolic reflections (estimation of affiliation)

Procedures for radargram examination for the needs of automated detection can be performed by either analyzing the full, dense radargram image or by analyzing a thresholded sparse version of it [2]. It is possible to apply unsupervised procedures (Hough transform, for instance) and supervised procedures (e.g., Artificial Neural Networks – ANN) if dense radargram is analyzed. According to the authors' best knowledge, all existing strategies for hyperbolic reflections detection involve application of algorithms that implement Hough transform [4], Wavelet transform [5], Radon transformation [6], standard algorithms for pattern recognition, such as Support Vector Machines (SVM) [7], or ANN [8]. ANNs are easiest to train after radargram simplification has been done by edge detection [9] or binarization [10]. ANNs can trained using signal processing statistical data descriptors [11], Welch power spectral density estimate of signal segments [12] or generated data sets which model EM waves propagation in specific conditions [13], [14].

The analysis of the papers related to these technologies yields the conclusion that the search through the dense radargram is very demanding in terms of time and computation resources and sensitive to noise and hyperbolic segments interference as well. Furthermore, it can be noticed that the majority of the procedures is directed towards the analysis of simplified radargram, which can be done in two ways [3]:

1. Simplification of dense radargram and extraction of the data from the hyperbolic reflection
2. Segregation of small two-dimensional sections from dense radargram (called segment of interest - SOI) and extraction of the data from the hyperbolic reflection

The procedure presented in this paper is based on the application of the procedure with segregation of segments of interests (SOI) from dense radargram. Such procedure retains maximum possible data resolution from the radargram in one or more segregated SOIs. When the data is extracted from SOI it is possible to apply each of aforementioned techniques for data extraction regardless of their complexity, because the amount of data is significantly reduced. The proposed procedure mostly resolves problems of automated hyperbolic reflections detection. When dense radargram is reduced to one or several SOIs, the amount of data is significantly decreased but data in each zone are unchanged.

When hyperbolic reflections are incomplete and noisy, there is a significant possibility of a complete or large loss of data during the simplification of dense radargram.

II. PROCEDURE DESCRIPTION

The proposed procedure is done in three steps.

Step 1: basic pre-processing of radargram in terms of time-zero offset and *.dzt to *.bmp conversion. Software tools provide functions for this conversion, both free software (e.g. "MatGPR" [21], or "Rad2bmp" [22]) and commercial software (e.g. "RADAN" [23]).

Step 2: detection of SOI in dense radargram using ANN, that is Cascade Object Detector trained on previously formed sets of samples which contain hyperbolic reflections (positive targets) and ones that do not contain hyperbolic reflections (negative targets).

Step 3: hyperbolic reflections segregation and data extraction (x,t) in the zone of interest using an edge detection technique.

With this input data set, estimation procedures are applied. Geometric and other characteristics of manmade objects are estimated, as well as EM wave's propagation velocity, which is described in subsection B.

1) Hyperbolic reflections detection (Step 2)

Cascade Object Detector (COD) is well known algorithm for machine learning. It is based on Viola-Jones learning algorithm [15]. First, the classifier has to learn to identify an object. It is done by training during which many positive (containing object - hyperbolic reflection) and negative (not containing object - without hyperbolic reflection) sample images are analyzed (Fig. 1). After the training has been completed the output classifier can be used for object recognition. In the process of detection, the algorithm divides the input images into many sub-images by moving a search window at multiple scales over it. The each sub-window is classified as object or no-object [16].

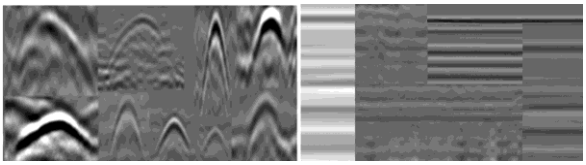


Figure 1. Example of positive (left) and negative (right) training samples

COD consists of stages with each phase representing a set of weak classifiers (Fig. 2). Weak ones are simple classifiers called 'part of decision'. Each stage is trained using the technique called 'boosting'. 'Boosting' provides the possibility to train highly precise classifier taking weighted average of decisions which are made using weak classifiers [16]. Each stage of the classifier marks a region defined by current location of the moving window as either positive or negative. Positive mark indicates that the object is found while negative mark indicates that there is no object of interest. If the mark is negative, classification in that region is over and detector moves the window to a next location. If the mark is positive, the classifier sends possibly positive region to next stage. Finally, a detector reports found object on the current location of the window, if the region is classified as positive in the final stage. Stages are designed to reject negative sample images as fast as possible. Presumption is that most of the windows do not contain object of interest. On the other hand, true positive objects are rare and it is worth to spend time for inspection of every single object. An object becomes true positive when positive sample image is correctly classified and false positive when negative sample image is misclassified as positive. An object becomes false negative when positive sample image is misclassified as negative. In order for algorithm to work correctly, each stage in the cascade must have low rate of false negative objects. If the object in the stage is mismarked as negative, classification is cancelled and it is not possible to correct the error. Moreover, each stage needs to have a high rate of false positive objects. Even if a classifier mismarks negative object as positive, the error can be corrected in next stages. Overall rate of false positive objects in cascade classifier is f^s, where f is the rate of false positives per stage in the range (0, 1) and s is the number of stages. Similarly, the overall rate of true positives is t^s, where t is the rate of true positives in the range (0, 1]. Hence, it can be noticed that adding of stages reduces overall false positives rate, but it reduces overall true positives rate as well [16].

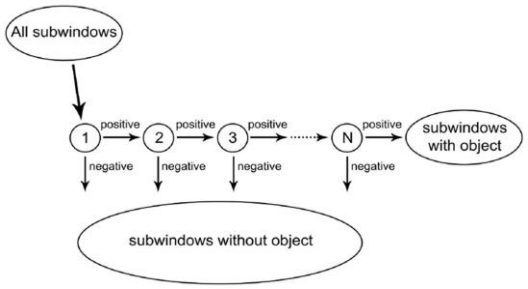


Figure 2. Schematic depiction of the detection cascade with N stages.

COD algorithm supports three training models:

- Haar-like features [16]
- LBP – Local Binary Patterns [17]
- HOG – Histograms of Oriented Gradients [18]

Each of these models has its advantages and disadvantages, so experimental evaluation of their applicability has been done, under the same conditions. The training set was comprised of more than 100 positive and more than 200 negative sample images. All used pieces of data are real (not generated) and collected in real conditions (urban and suburban area), not on test-fields (which have strictly defined parameters and acquisition conditions). The radargrams are collected using 200, 400 and 900MHz antennas, with several variations in terms of type of soil, homogeneity and volumetric moisture content.

Fig. 3 represents comparative analysis of the results of the mentioned models application. In the Fig. 3, the radargram is shown as *.bmp with time-zero offset removed. The radargram is formed on a soil with low volumetric moisture content with two pipes of large diameter (500 and 350mm) in it.

The key parameters of the experiment were training speed and accuracy (number of 'false' targets). Results yielded following conclusions:

1. Haar-like features – longest training, medium number of 'false' targets that are always present, which indicates higher sensitivity to interference.
2. LBP – shortest training, highest number of identified 'false' targets that are always present which indicates higher sensitivity to interference.
3. HOG – medium time for training, lowest number of 'false' targets identified, in most cases none.

Fig. 3a (Haar model) shows that in represented example 3 'false' objects are registered, while the time for training was 238 seconds. Fig. 3b (LBP) shows that, in represented example, 6 'false' objects are registered,

while the time for training was 7 seconds. Finally, Fig. 3c (HOG) shows that not a single 'false' object was registered in this example, while the time for training was 52 seconds.

Number of detector stages has to be defined relative to a rate of 'false' positive targets per stage: lower rate yields smaller number of stages, and vice versa. Generaly, it is better to have bigger number of simple stages with since the rate of 'false' positive targets decreases exponentionaly. Number of COD stages that produced best results in experiments was around 20.

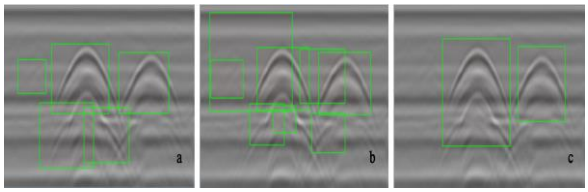


Figure 3. Results of COD algorithm trained on three data set models.

Processing of the single radargram lasts from one to several seconds depending on the length and complexity of radargram and computers processing power.

2) Hyperbolic reflections segregation and data extraction (Step 3)

Edge detection is an image processing technique, designed to recognize the edges of the object within the image [19]. Function processes a grayscale image and returns binary image with the same dimensions as the grayscale source. Binary image has 'ones' where the edge of the object is found and 'zeroes' elsewhere. Edges match significant local changes of the intensity in the image. Edge is a set of connected pixels that are on the borderline between two regions. Intensity changes are caused by different physical changes, including color, texture, reflection and shadows. Edge detection on noisy images is very difficult, since both the noise and the edges contain high-frequency components; moreover not every edge contains gradual changes of intensity [5].

Edge detection is done using "canny" operator that finds the edges by local maximum of image (radargram) gradient, which is calculated using Gauss filter [19]. The method uses two threshold values in order to detect strong and weak edges, and includes weak edges into output result only if they are connected with strong edges. The probability that the function with canny parameter will be misguided by the noise is lower, while the probability that true weak edges will be detected is higher. Application of functions for edge detection results in smaller amount of data that needs to be processed and makes this system functional in real time [12].

The result of edge detection function is boundary area (Fig. 4a). Overlapping of boundary area with the SOI, all the pixels within boundary area are multiplied by 1, while pixels outside the boundary area are multiplied by 0. That extracts only pixels within boundary area that belong to hyperbolic reflection only. Fig. 4b represents detection of maximum of pixel intensity along columns in order to ensure that only strongest reflection is saved.

These are the maximums that appear for the first time, i.e. if in one column the same maximum appears several times, the first one is taken into account.

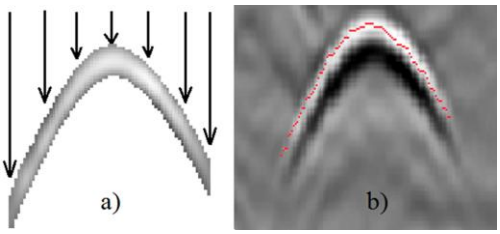


Figure 4. Hyperbolic reflection data segregation a) and finding maximum values in columns b).

3) Method to simultaneously estimate the radius of a cylindrical object (R) and the wave propagation velocity (v) - short review

Results produced by automated hyperbolic reflections procedure are used to estimate radius of a cylindrical object and the wave propagation velocity simultaneously. Considering these procedures together, they form a complete solution for automated detection of cylindrical underground objects and estimation of their parameters. Our estimation method is not the main focus of this paper so it is given here in a short review. Further details can be found in [1].

Equation (1) defines non-linear estimation model with 4 variables: x_0, t_0, R and v where (x_0, t_0) are the apex coordinates of the hyperbola optimally fitted through raw data [1].

$$\frac{\left(t+\frac{2\cdot R}{v}\right)^2}{\left(t_0+\frac{2\cdot R}{v}\right)^2}-\frac{(x-x_0)^2}{\left(\frac{v}{2}\cdot t_0+R\right)^2}=1 \qquad (1)$$

(x, t) – extracted point coordinates.

Step 1: (x_0, t_0) is estimated using a modified Levenberg–Marquardt method. The applied algorithm is robust and was adapted to solve nonlinear problems using the least squares method. The basic task of the first step is to decrease the number of correlations between the estimated parameters, i.e., to reduce the problem dimensionality from four to two correlated parameters. The estimated values (x_0, t_0) are the input data for the next step.

Step 2: estimation of boundary speed v_0. Since v is unknown (or approximately known, which is insufficient to accurately estimate R), an additional condition to simultaneously estimate v and R is defined as the velocity range [v_{max}–v_{min}]. The boundary velocity is estimated iteratively by varying v. The value closest to satisfying the condition $R\approx 0$ is accepted as v_0. A propagation velocity higher than v_0 does not make sense, because it produces negative values of R. With values known from the previous step nonlinear model (1) enables determination of unique estimation of v_0.

Step 3: simultaneous estimation of v and R. Since v is varied in the interval $[v_{max}–v_{min}]$ forming a criterion to choose v enables the selection of optimally estimated v from the set of possible values. For the selection criterion the *foo (First Order Optimality Criterion)* is chosen.

III. Experimental Results

Verification of the method was done on a number of radargrams (more than 100 typical cases). All of them contain real field data and geometry of underground objects is known (ground truth data!). Two typical radargrams are selected for discussion in this paper. The first one has low noise ratio and interference, while the other represents the opposite case.

The radargram, 6.82m long, shown in Fig. 5, is formed using 400MHz antenna, with 1 scan/cm horizontal and 512smp/scan vertical resolution. Corresponding bitmap has dimensions of 682x512 pixels. Detection of two hyperbolic reflections is correct. 203 points are extracted for the detected hyperbola on the left side (500mm diameter pipe), while 166 points are extracted for the hyperbola on the right side (350mm diameter pipe). This example contains several hyperbolic reflections (two) which are relatively close, and to a certain degree interfere with each other.

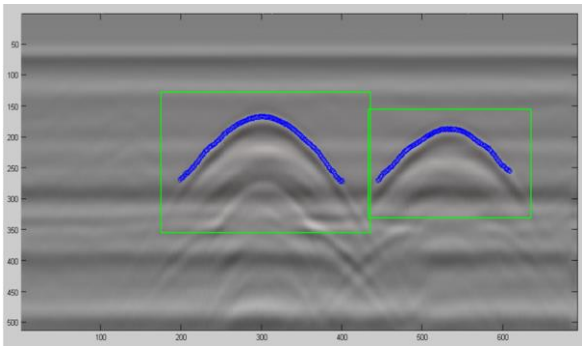


Figure 5. Original radargram and result of detection.

The radargram in Fig. 6 is 461 scan long. Five underground tanks were scanned, with vertical resolution of 512 smp/scan. Since tanks are of larger dimensions and adjacent, the interference between hyperbolic reflections is evident (marked with arrows on Fig. 6). The interference resulted with ‘false’ hyperbolic reflections. Five ‘true’ hyperbolic reflections were detected along with one ‘false’.

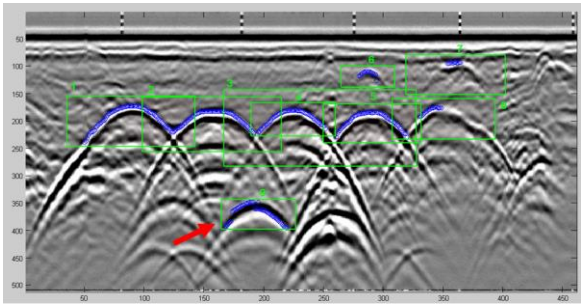


Figure 6. Original radargram and result of detection.

Detection of hyperbolic reflections proved itself as very robust to noise and incomplete hyperbolas, while not that robust to interference (it has to be emphasized that interference is not that common in real application). Edge detection also proved to be very robust, but radargrams with low signal intensity sometimes posed a problem. In some cases applying a gain to a certain degree enabled edge detection.

IV. Conclusion

In this paper, new procedure for automated detection of hyperbolic reflections and data extraction from radargrams was proposed. Procedure was tested on a number examples containing real field data and some of experimental results were shown in this paper. Experiments showed that the procedure is robust to various types of media surrounding the objects, incomplete or noisy hyperbolic reflections and that it produces satisfactory results for the needs of underground utility detection. Robustness to interference of neighboring hyperbolic reflections can be improved by adding expert knowledge to the procedure. Along with our existing parameter estimation method it represents the complete solution for underground utility characterization.

Acknowledgment

Some results represented in this paper are obtained through the project "Modeling the state and the structure of slope processes, using GNSS, TLS and GPR", founded by Ministry of science and education, project number TR 37017.

References

[1] Ristic, D. Petrovacki and M. Govedarica, “A New Method to Simultaneously Estimate the Radius of a Cylindrical Object and the Wave Propagation Velocity from GPR Data,” *Computers & Geosciences*, vol. *35*, pp. *1620-1630*, Aug., 2009.

[2] R. Janning, A. Busche, T. Horváth and L. Schmidt-Thieme, “Buried pipe localization using an iterative geometric clustering on GPR data”, *Artificial Intelligence Review*, vol. *42*, no. *3*, pp. *403-425*, Oct., 2014.

[3] S. Birkenfeld, "Automatic detection of reflexion hyperbolas in GPR data with neural networks", *in Proc. World Automation Congress (WAC 2010)*, Sept. 2010, Kobe, Japan, pp. *1189-1194*.

[4] G. Borgioli, L. Capineri, P. Falorni, S. Matucci, C. Windsor, 2008. “The detection of buried pipes from time-of-flight radar data,” *IEEE Trans. Geosci. Remote Sens.*, vol. *46*, no. *8*, pp. *2254–2266*, Aug., 2008.

[5] H. Zhou, M. Tian, X. Chen, “Feature extraction and classification of echo signal of ground penetrating radar,” *Wuhan University Journal of Natural Sciences*, vol. *10*, no. *6*, pp. *1009-1012*, Nov., 2005.

[6] A. Dell’Acqua, A. Sarti, S. Tubaro, L. Zanzi, “Detection of linear objects in GPR data”, *Signal Processing,* vol. *84*, no. *4*, pp. *785-799*, Apr., 2004.

[7] E. Passoli, F. Melgani, F. Donelli, "Automatic Analysis of GPR Images: A Pattern-Recognition Approach," *IEEE Trans. Geosci. Remote Sens.*, vol. *47*, no. *7*, pp. *2206–2217*, July, 2009.

[8] H. S. Youn, C. C. Chen, “Automatic GPR target detection and clutter reduction using neural network,” *in Proc SPIE 4758, 9th International Conference on Ground Penetrating Radar*, S. Barbara, 2002.

[9] M. R. Shaw, S. G. Millard, T. C. K. Molyneaux, M. J. Taylor, J. H. Bungey, “Location of steel reinforcement in concrete using ground penetrating radar and neural networks,” *NDT & E International, vol. 38*, no. *3*, pp. *203-212*, Apr., 2005.

[10] P. Gamba, S. Lossani, “Neural detection of pipe signatures in ground penetrating radar images,” *IEEE Trans. Geosci. Remote Sens.,* vol. *38*, no. *2*, pp. *790-797*, Mar., 2000.

[11] S. Shihab, W. Al-Nuaimy, Y. Huang, A. Eriksen, “Neural network target identifier based on statistical features of GPR signals,” *in Proc SPIE 4758, 9th International Conference on Ground Penetrating Radar,* S. Barbara, 2002.

[12] W. Al-Nuaimy, Y. Huang, M. T. C. Fang, V. T. Nguyen, A. Erikson, “Automatic detection of buried utilities and solid objects with GPR using neural networks and pattern recognition,” *Journal of Applied Geophysics,* vol. 43, no. *2-4*, pp. *157-165*, 2000.

[13] F. Frezza, L. Pajewski, C. Ponti, G. Schettini, N. Tedeschi, “Cylindrical-Wave Approach for electromagnetic scattering by subsurface metallic targets in a lossy medium,” *Journal of Applied Geophysics*, vol. *97*, pp. *55-59*, Oct., 2013.

[14] A. Giannopoulos, “Modelling ground penetrating radar by GprMax,” *Construction Building Mater.*, vol. *19*, no. *10*, pp. *755–762*, Dec. 2005.

[15] Viola, P., Jones, M., “Rapid object detection using a boosted cascade of simple features,” *In: Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, IEEE Computer Society, 2001, pp. *511-518*.

[16] C. Mass, J. Schmalzl, "Using pattern recognition to automatically localize reflection hyperbolas in data from ground penetrating radar", *Computers & Geosciences*, vol. *58*, pp. *116-125*, Aug., 2013.

[17] T. Ahonen, A. Hadid, M. Pietikainen, "Face Recognition with Local Binary Patterns", *in Proc. Computer vision - ECCV 2004, Lecture Notes in Computer Science*, vol. 3021, pp. 469–481.

[18] N. Dalal, B. Triggs, "Histograms of Oriented Gradients for Human Detection", *in Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR '05)*, Jun 2005, San Diego, USA, pp. *886-893*.

[19] J. Canny, "A computational approach to edge detection", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. *8*, no *6*, pp. *679-698*, 1986.

[20] http://users.uoa.gr/~atzanis/matgpr/matgpr.html

[21] http://www.geophysical.com/softwareutilities.htm

[22] http://www.geophysical.com/software.htm

Orchestrating Music Queries via the Semantic Web

Milos Vukicevic, John Galletly
American University in Bulgaria
Blagoevgrad 2700
Bulgaria
+359 73 888 466
milossmi@gmail.com, jgalletly@aubg.bg

***Abstract* - This paper describes the design and implementation of a Semantic Web application that allows queries and inferences to be made on a music knowledge base using Semantic Web technologies such as RDF, OWL and SPARQL. Additionally, the paper explains how these technologies were blended together to develop the application that illustrates the principles of the Semantic Web.**

I. INTRODUCTION

The Semantic Web has been heralded by the W3C as the future web – a web that relies heavily on the software implementation of knowledge bases and inference mechanisms [1]. The Semantic Web has several standards recommended by the W3C with, currently, varying levels of functionality and usability [2]. It is still heavily under development and evolution, with the latest standard coming out in 2014. The Semantic Web software stack [3] is illustrated in Figure 1

The application described in this paper is a Semantic Web application that allows music queries and inferences to be made on a music knowledge base. The word "knowledge" is important – a traditional database approach would not give the breadth and scope for queries and inferences that a knowledge base (expressed as an RDF ontology) would.

The design and implementation of a fully-fledged music ontology was beyond the scope of this work. Rather than rely on a large, ready-built music ontology (e.g. mucicontology.com), the application described here was developed with a much narrower ontology, namely one for rock music and bands. But, given enough time and effort, the design described here could be extended to cover different types of music and artists. Information about artists, tracks, etc. in this ontology, is represented as RDF statements.

The use of the Semantic Web technologies in the music industry is not new. For example, the BBC's Music Project is an effort by the BBC to build semantically-linked and annotated web pages about artists and singers whose songs are played on BBC radio stations [4].

II. DEVELOPMENT ENVIRONMENT

Apache Jena [5], in conjunction with the Eclipse IDE, was used as the basic programming environment. Jena is a Java-based API for Semantic Web development. It provides extensive Java libraries for handling RDF, OWL and SPARQL in line with the published W3C recommendations. Jena includes a rule-based inference engine to perform reasoning based on OWL and RDFS ontologies, and a variety of storage strategies to store RDF triples.

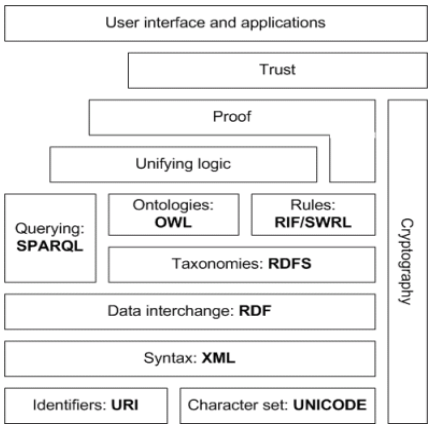


Figure 1 – Semantic Web software stack

The Stanford Protégé ontology editor [6] was used to build the application's ontology, as this editor provides an easy-to-use environment for developing ontologies. The ontology was developed using OWL Description Logic (OWL DL). Once built, the ontology was then loaded into Jena. Jena's generic inference mechanism was used to make inferences between the ontology classes. Additionally, the Jena SPARQL query engine allows for expressive SPARQL queries. However, it does not contain the full implementation of SPARQL as it is envisioned by the W3C. It is impossible for the user to create resources or add properties, only to search for already existing graph patterns.

III. DESIGN

A. Design Overview

From the outset, the design of the software was made to be scalable, and is essentially developed with an MVC pattern, where the GUI is the View, the ontology is the Model, and the Jena inference engine and SPARQL query engine are the Controller [7, 8].

Basically, the ontology is used as the basis for executing SPARQL queries, and making inferences using the Jena inference mechanism. The ontology had to be extensible in terms of having the ability of adding new ontologies to it and expanding the ontology itself, while also providing a scalable ground for adding new instances of ontology classes, etc.

With the limits imposed by the current standards, and by the architecture of Jena, the SPARQL queries had to be created programmatically to fit the ontology. The queries had to be designed in such a way that they would operate with the architecture of the ontology in question, making full use of the

data and logic provided by it. Moreover, the application's SPARQL interface had to be designed in such a way that it would handle additions to the ontology, and ensure that the program would still work correctly, even with these additions.

The GUI is the front end, and is able to accept four types of queries: queries for a track, album, artist or band. The results are shown in three screen text panels, one containing basic data inferences, the other containing basic relationship inferences (such as Artist X plays in Band Y), and advanced inferences linking independent nodes together semantically (Figure 2).

As the application was designed with scalability in mind, adding more query types to the list would not be too difficult, but the ontology, as is designed currently, requires no further subtypes.

The ontology consists of several top-level classes. These classes all have instances of themselves, in some case multiple instances. The semantic web allows for a dynamic addition of other instances of these classes, even of other classes. The ontology class design can be seen in the Figure 2.

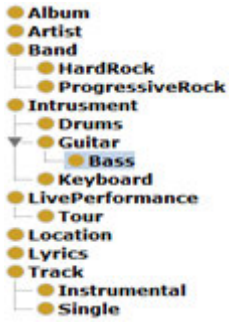


Figure 2 – Ontology top-level classes

The relationships between classes are defined using object properties. Figure 3 is a list of all object properties in the ontology.

Object properties act as predicates between individuals but no literals. Predicates for literals are data object properties and they are illustrated in Figure 4.

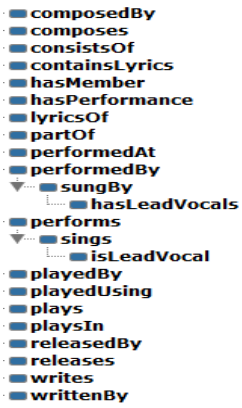


Figure 3 – Ontology object properties

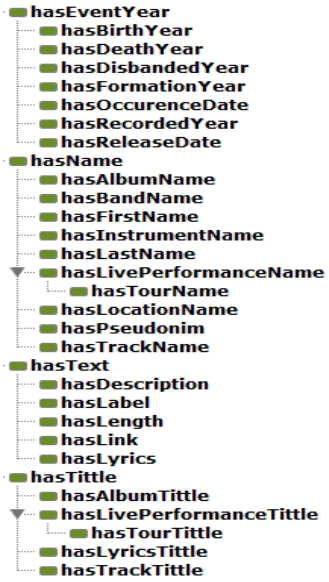


Figure 4 – Ontology data object properties

Figure 5 shows the GUI, the SPARQL engine and ontology packages with their dependencies. The GUI relies on the SPARQL engine to populate it with data. The GUI package has various elements and functions that allow it to display the data properly and also capture button click events. The SPARQL Engine has all the necessary data structures and functions to run queries, process them, and perform advanced inference.

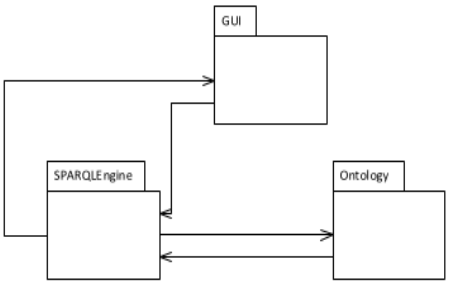


Figure 5 – UML package diagram

Figure 6 illustrates the deployment of the packages.

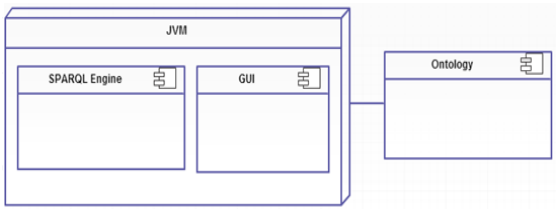


Figure 6 – UML deployment diagram

B. Design Details

The ontology is stored as a separate file using the RDF/XML standard, and it is then imported into the Java application using the Jena ModelFactory pattern. The ontology contains the ontological specifications, i.e. all the classes, object properties and data properties available for the specific ontology, along with all the individual instances of the classes and their predicates. It is designed so it is extensible, i.e. new classes can be added, as well

as new ontologies, and it is also scalable, i.e. new individuals can be added without negatively impacting the execution of the entire software solution.

The Jena inference and SPARQL query engines operate on the ontology. After being loaded into the application, the inference engine is run on the ontology and an "inferred" model is created. This is an extended, in-memory version of the ontology, providing advanced inferences about the classes and properties. This inferred model is then used as a basis for various SPARQL queries.

The Semantic Web can essentially allow for an extremely large amount of semantic queries (such as "Who played the guitar at concert X?") and therefore needs some kind of query parsing or translation mechanism to allow the application to "understand" what exactly it is that the user is looking for. This is programmatically a challenge in its own right, and there was not enough time to implement such an input parser. However, having the ontology in mind, the interface to the SPARQL engine was constructed in such a way that it is able to return complex inferences from the ontology itself for a particular set of search strings.

While the user is able to perform basic semantic querying, the SPARQL interface takes the particular query of the user and retrieves additional advanced semantic inferences about the particular object the user is looking for. This was accomplished by generating inferred assertions using Protégé's inference engine operating on the ontology, as the ontology was built.

C. The SPARQL Interface

This is the "heart" of the application. This part revolves around reading the user input, and then trying to match it to a list of all albums, artists, bands, or tracks, depending on what the user has selected. If there is a match, this is then processed and a query is created that can be run against the ontology. This module is also responsible for loading the ontology, creating an inferred model using the Jena inference mechanism and then running queries on the inferred (in-memory) model.

There are several operations that need to be performed before doing so. The most straightforward function is the URI dereferencing. All entities in the ontology have a URI namespace prefix. For example, the Pink_Floyd instance of the class Band always has the entire namespace prefixed to it, so it would be:

http://www.semanticweb.org/milos/ontologies/2014/3/music #Pink_Floyd

Before an entity can be searched for, the namespace must be removed.

Similarly, there is an algorithm that prepares an entity for output based on its type, i.e. Artist, Band, Album, Track, etc. This turns Pink_Floyd into "Pink Floyd," for example.

The SPARQL query is built functionally. The example below demonstrates the SPARQL query interface. It takes in the query type, which would be SELECT in most cases, the string pattern which is the subject of selection, the subject of the WHERE clause, the predicate of the WHERE clause, and the object of the WHERE clause. This returns a distinct result set which is passed onto a globally declared variable called resultArray. The resultArray is an ArrayList of type string that stores all the information a SELECT query returns. The main application then deals with the returned data in some way.

```
// Generic Query creation engine
// Result of Query passed to global variable resultArray
private static void runQuery(String queryType, String pattern, String subject, String predicate, String object)
{
        // Clear Result Array
        resultArray.clear();

        StringBuffer queryStr = new StringBuffer();
        // Establish Prefixes
        //Set default Name space first

        queryStr.append("PREFIX rdf" + ": <" + "http://www.w3.org/1999/02/22-rdf-syntax-ns#" + "> ");
        queryStr.append("PREFIX owl" + ": <" + "http://www.w3.org/2002/07/owl#" + "> ");
        queryStr.append("PREFIX xsd" + ": <" + "http://www.w3.org/2000/01/rdf-schema#" + "> ");
        queryStr.append("PREFIX music" + ": <" + "http://www.semanticweb.org/milos/ontologies/2014/3/music#" + "> ");
```

The code below illustrates the second part of query execution, where the query is formulated and executed using the functions provided by the Jena ModelFactory. Both resources and literals are retrieved in this fashion with proper formulation of queries. However, the advanced inference relies on a programmatically use of several query calls, relating individuals that are not usually directly related - more on this in the implementation section.

```
//Now add query
String queryRequest = queryType + pattern + " WHERE { " + subject +" " + predicate + " " + object + " }";
queryStr.append(queryRequest);
Query query = QueryFactory.create(queryStr.toString());
QueryExecution qexec = QueryExecutionFactory.create(query, inferredModel);
try {
ResultSet response = qexec.execSelect();

while( response.hasNext())
{
    QuerySolution soln = response.nextSolution();
    RDFNode name = soln.get( pattern);
    if( name != null )
    {
        resultArray.add(name.toString());
    }
}
} finally { qexec.close();}
}
```

D. The GUI

The GUI accepts and parses user input data, and displays the results of the query and the relevant basic and advanced inferences on the screen. The front end was simplified to provide scope-limited queries, in the sense that the user could query for specific information while the inferences were prebuilt into the ontology itself. That is to say, the user could query to find an artist, and the artist would be found, while advanced inferences about the artist are displayed in the information boxes. The screen itself is split into five panels (Figure 7). The top panel is the search box and it does not change. It contains a combo box allowing the user to select the type of query he/she wants to perform (Find artist, album, track or band), a textbox for the actual query string, and a button to initiate the query. The other four panels are used to display the data retrieved. The first and top left panel of the four displays the relevant image associated with the query. The second panel contains basic data inferences, such as data properties. The third panel contains subject assertions, i.e. the correlation of the searched subject with all the other subjects that the searched subject is immediately connected to in terms of the semantic graph. The fourth panel contains the advanced queries, linking multiple nodes that are not directly correlated, or performing operations on existing data.

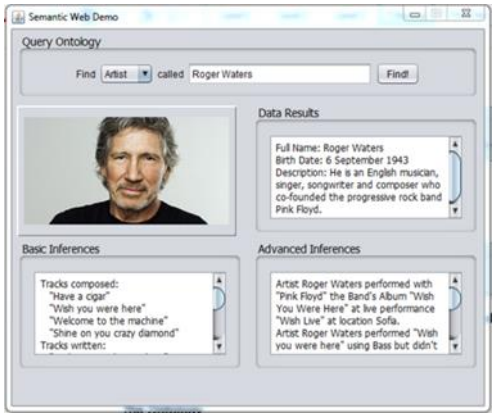


Figure 7 – Finding an artist

Figure 7 and Figure 8 illustrate the functionality for finding a particular named artist and a particular named track.

Figure 8 – Finding a track

The last element of the GUI is the "Play Track" button which is hidden at the bottom of the page and is only displayed once a track is searched for. If clicked, it will open a new frame which opens a relevant YouTube link to the track in question (Figure 9).

Figure 9 – YouTube link frame

E. Semantic Web Implementation

RDF is the Semantic Web notation for modelling application domain information. The information is actually represented in the form of a graph database. "Pieces" of information are represented in the form of assertions called statements and each statement is made out of three parts (or triples): a subject, a predicate and an object.

Each subject and object represents either a resource or literal, while the predicate illustrates the relationship between subjects, objects, and literals.

While RDF allows the description of domain resources and relationships using domain vocabularies, it does not support semantics. RDF Schema (RDFS) is an extension to RDF that allows the description of semantics in terms of classes, instances of classes, hierarchies, etc. RDFS allows the creation of disjoint properties, the specifications of types, domains and ranges, as well as indicating the type of properties in terms of their being functional, reflexive, and transitive, etc. This permits a simple implementation of semantics that is used as the basis for the powerful Web Ontology Language, OWL.

W3C developed OWL as a standardized way of expressing higher-level data semantics in Semantic Web applications. Like RDF and RDFS, OWL has an XML-based syntax. It comprises several sections. The first section is a header section where appropriate OWL namespaces are referenced. This section is followed by class declarations, object properties and data properties. After these declarations comes the individual specifications section, which declares instances of the classes and uses the aforementioned object properties to link different individuals together, while the data properties are used to link individuals with literals.

The following is an example of an OWL expression used to declare an individual of class Artist.

SPARQL is in many ways similar to SQL but it is for the Semantic Web. For example, the SELECT command specifies the result set and its name, while FROM clause indicates which file or SPARQL endpoint will be queried for the result. A WHERE statement specifies the graph pattern to be searched for, while ORDER can be used for data result formatting.

```
<NamedIndividual rdf:about="&music;Nick_Mason">
    <rdf:type rdf:resource="&music;Artist"/>
    <music:hasBirthYear>27 January 1944</music:hasBirthYear>
    <music:hasLastName>Mason</music:hasLastName>
    <music:hasDescription>He is an English musician and composer, best known as the drummer of Pink Floyd. </music:hasDescription>
    <music:hasFirstName>Nick</music:hasFirstName>
    <music:plays rdf:resource="&music;Drums"/>
    <music:playsIn rdf:resource="&music;Pink_Floyd"/>
    <music:performs rdf:resource="&music;have_a_cigar"/>
    <music:performs rdf:resource="&music;shine_on_you_crazy_diamond"/>
    <music:performs rdf:resource="&music;welcome_to_the_machine"/>
    <music:performs rdf:resource="&music;wish_you_were_here"/>
</NamedIndividual>
```

The above diagram illustrates a simple SPARQL query. The PREFIX statements define various namespaces, with "music" being the namespace for the ontology. The query SELECTs a subject (?Track) that fits into the graph pattern subject-predicate-object (music:Roger_Waters - music:composes - ?Track). This query will return a list of all Tracks composed by artist Roger_Waters. The program iterates through all top-level entity object properties and appends them to the basic inferences text area.

The advanced inferences rely on multiple levels of connection and making further inferences. For example, the inference "Roger Waters has played with Pink Floyd, the album "Wish You Were Here" at the concert "Wish Live" in Sofia", is deduced in this way. A number of special algorithms were developed to make these inferences for artists, tracks, bands, etc. For example, the algorithm for the above inference is

1. An artist is selected.
2. A list of band(s) is found through basic inference.
3. A list of all matched bands is looped through, and a list of all albums is found through basic inference through the Band, connecting the Artist with the album (no direct connection).
4. A list of all matched albums is looped through, and a list of all live performances is found through basic inference through the Album, connecting the Artist with the live performance and the band with the live performance. (no direct connection)
5. A list of all matched live performances is looped through, and a list of all locations where the performances were held is found, connecting the album, artist, and band with the locations.
6. A list of all locations where the live performance was held is looped through, connecting the Album with the location, the Band with the location, and the Artist with the location, and the results are printed recursively at this point from the location back to the Artist. (no direct connection)

IV. CONCLUSION

The design and implementation of a Semantic Web application, that handles music queries for a limited domain, has been described. In principle, the application could be further extended as a comprehensive, semantically-organized music knowledge base with support for all types and genres of music, ranging from modern rock and roll and pop, to classical music and classical pieces of music.

REFERENCES

[1] W3C Semantic Web: http://www.w3.org/standards/semanticweb/

[2] W3C Semantic Web Standards: http://www.w3.org/standards/semanticweb/

[3] Wikipedia: Semantic Web: http://en.wikipedia.org/wiki/Semantic_Web

[4] BBC Music Project

http://readwrite.com/2009/01/21/bbcs_semantic_music_project

[5] Apache Jena Documentation: https://jena.apache.org/documentation/

[6] Protégé Wiki: http://protegewiki.stanford.edu/wiki/Main_Page

[7] Strategies for Building Semantic Web Applications: http://notes.3kbo.com/sparql

[8] Semantic Web Programming: https://code.google.com/p/ia1213/downloads/detail?name=semantic-web-programming.9780470418017.47881.pdf

REPORTING SYSTEM FOR MOBILE

Dr. Szilveszter Pletl[1], Gabor Pletl[2], Regina Seres[3]
Institute of Informatics, University of Szeged, Hungary, Business informatics, University of Szeged, Hungary[2], Business informatics, University of Szeged, Hungary[3]

Abstract -*The system described in this paper is a real time and interactive reporting system with information visualization. Currently and in the near future the services that are able to reach real time data, visualize it and offer a clear view of the presented information, are becoming more and more essential in the enterprise sector. These solutions are called "one button" solutions. The information needs to be available to the managers and company leaders from anywhere in the right format. The presented system is the core of a decision support system.*

Keywords–*security, information systems, reporting system, Future Internet*

1. INTRODUCTION

The system described in this paper is a real time and interactive reporting system with information visualization. At first, an overview will be given about the system. In the overview section the outline of the software components will be presented and the processing workflow described. The fourth section offers the classification of the product, and the mentioning of the possibilities regarding further development. The fifth section contains the communication process between the system components. This description provides an overview of the process though it does not detail the exact protocol of the communication. A new generation framework name FIWARE[1] was used. In the sixth section the authors answer why it is important to use the Future Internet, which is a general term for research activities on new architectures for the Internet. Data security is described in the seventh section where details are given about the benefits of this system and specifications of the internal security solutions. At this point the disadvantage of the system will also be referred to. In the network security section focus will be on the communication security. The test section of the paper shows the response time according to the amount of records processed by the enterprise server. The final section provides a summary of the benefits and disadvantages of the present system and finally conclusions are drawn based on the experiences gained.

2. GOAL OF THIS RESEARCH

We think that the internet nowadays is very differ from the internet in the past, but we often use technologies that are based on the early internet. The bandwidth, reliability and other qualities made it possible to take the internet technology on higher level. For this reason we considered it important to satisfy the future internet concept. The main goal of the project was to design a safe model for the sensitive communication, data processing by conventional means, join different data sources, visualize the results and try to save as much time as possible. First, we had to figure out a powerful system model for the communication. It is a three-component solution which is described in the fifth section.

3. OVERVIEW

The presented system satisfies the concept of the Future Internet through the FIWARE architecture that will be detailed in this paper. The system has three components: a public server, corporate server and a mobile client. The public server handles the connection between the corporate server and the mobile client. It has logging, user managing and subscription handling tasks, as well. The corporate server is private and every company that uses the system needs to install its own corporate server. It provides a web interface for administrators to build the charts and handle the user eligibilities. This is the source of the chart menu in the mobile client. The mobile client is a lightweight software. It needs to be, because it is very risky to store enterprise data on the client side. The mobile client has two main tasks: authentication and visualization. In this paper further details will be given of these components, highlighting the importance of the Future Internet and explaining how this present solution is implemented.

4. PRODUCT POSITIONING

Interactive software that promotes the effective functioning of the groups or communities in a form of business process podcast, tracking and other functionalities is a decision support system. The core task of the corporate decision-making system is the proper service of the information from the raw data. There is more than one type of the DSS, but the authors will not go into details because it is not the purpose of this paper. The following section will present a summary of the peculiarities of the DSS to facilitate the position of the created system.

Today there is no "big" corporation that could work without a DSS. The role of the corporate decision support systems to facilitate the decision taken by the conclusions from the data processed by them. The data-driven decision support system has six potential present levels are distinguished. The less valuable raw data towards the useful information the following processes belong to the order:

- management of databases
- management data warehouse
- data extraction and purification
- data mining
- reporting and visualization

- decision support functionalities

The data processing and visualization is located at the level directly below the intelligent decision-making, so in this case we can talk about a reporting system. The system is able to be imporved by modules with decision support and in this way it can be a full valued DSS.

5. WORKFLOW

This chapter will outline the three component system so the reader is presented an overall picture of the operation of the system. Figure 1 below shows the system and how the components connected. This figure is detailed in this section of the paper. The public server, the corporation server and the client collaboration need to achieve the proper operation of all three elements and it is necessary to reach each other's network as well. The network access is not always clear, in any case, the SSL 3.0 standard is used according to the provisions. The following section describes the operation of the system by the use of sequence.

In the first step, the user can access the public server via the website where the private server can be downloaded. Before downloading, the new user is required to enter some personal information therefore the user can be identified later. The server generates a unique identifier and provides a download link. The corporate server installation requires Tomcat and MySQL server. The Tomcat helps not only carrying the administrative activities, but clients are communicating via this interface, as well. After the installation, the server is ready to be registered with the unique ID which the user received from the public server. After the product is registered the public server stores the IP address of the private (corporate) server, which then can be changed by the administrator. After the corporate server is registered, you can also register the clients. Following the downloading and filling out the registration formula with the unique key, an unlimited number of mobile clients can be registered on the public server and the corporate server. At first the registration request is received by the public server and if the unique id is valid it responds with the corporate servers IP address and makes some database updates to store the new client data. After this the client makes a registration request for the corporate server with the given username and password. Then the server permits the request and the new client is ready to use the service.

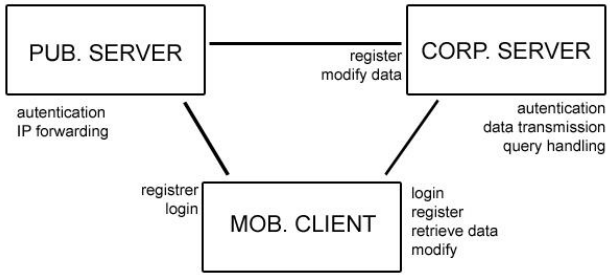


Figure 1. System communication

Once the client is registered, the system is ready for them to log in. Each time the client wants to log in to the server it performs a request to the public server. If the public server authenticates the client and identifies the IP address of the corresponding private server then sends it to the client who, in the second phase of the authentication, indicates a login request to its corporate server. Following the end of the authentication procedure, the public server does not play a role in communication anymore. It is a preferred solution mainly for security reasons, because sensitive corporate data is transmitted P2P bypassing the third party access.

6. FRAMEWORK

During the construction of the software design efforts were directed to satisfy the Future Internet intention. Future Internet is a general term for research activities on new architectures for the Internet. Non-technical aspects of a Future Internet span large areas such as socio-economics, [2] business and environmental issues. The Organisation for Economic Co-operation and Development held a conference called "Shaping Policies for a Digital World" in 2008. It proposed activities such as publishing recommendations for the future of the Internet economy.[3] Research areas that could be seen as components of a Future Internet include network management,[4][5][6] network virtualization, and treating any kind of information as objects, independent of their storage or location. Elements of cloud computing blended into the notion of Future Internet, leading to the concept of cloud networking.

The present goal was to take part in the spread of the Future Internet. One of the Future Internet projects is FIWARE, which is very popular in the European Union. FIWARE seeks to provide a truly open, public and royalty-free architecture and a set of open specifications that allow developers, service providers, enterprises and other organizations to develop products that satisfy their needs while still being open and innovative. FIWARE will dramatically increase Europe's Information and Communications Technology competitiveness by introducing an innovative infrastructure that enables cost-effective creation and delivery of versatile digital services, high-quality of service and security guarantees. The project offers open APIs that allow one to avoid coming tied to a specific vendor, therefore protecting one's investment. It provides a powerful foundation for the Future Internet and cultivating a sustainable ecosystem. The FIWARE project is under development and it will offer frameworks in different areas of development. The authors implemented the FIWARE security framework that provides the mechanisms which ensure that the delivery and usage of services is trustworthy and meets security and privacy requirements.

7. DATA SECURITY

The public server is available for everyone maintained by a third party and it has only a user management function. This means that the public server is never involved in meaningful communication between the client and the enterprise server, thereby ensuring the full independence of the data. Unfortunately, this decision comes at a price, because in addition to the minimal system requirements other background information about the hardware is not known. The cloud-based reporting systems can operate more efficiently because of their ability to provide distributed systems to process the data from the companies. They are able to process much faster thanks to the big data methods.

The data in the database is stored in encrypted form. This means that the data used during the tasks must be decrypted during the processing and encrypt when it was stored. The Mcrypt is a PHP library that implements a variety of data encryption algorithm and it was used for data encryption. After choosing the algorithm, the data is stored encrypted with the key that is constant.

If the client wants to read data or data line from the database, the encryption key and the encryption algorithm decoder pair decrypt the selected data. The data, which is stored only for coordination, typically passwords encrypted with hash algorithm. These algorithms cannot be decrypted programmatically; or at the very least, it takes irrationally long time to decrypt them. The hash functions' disadvantage is that collision generation can "decrypt" the data, which means that a given text's hash could be the same as the hash of another text. Therefore the hash function must be chosen very carefully. SHA-512 could be an appropriate choice, so user passwords were stored using this method [7] [8] [9]. User specific information like password and subscription data are stored on the client side, as well, which are not lost after the application is closed. The user specific information includes the username, password and product ID that is proved by the central public server each time when login is requested. There are more sensitive data storage solutions that can be used on the client side.

8. NETWORK SECURITY

From the network security point of view the flexibility of the system is a useful property. This means that the client server communication is fully customizable by the enterprise administrator of the service. Contrary to other products on the market it is not necessary to assign a priority gateway for the public server, because after the registration of the corporate server the administrator's decision, when connect to a public server. The private corporate client server communication allows the use of specific configuration settings. As already mentioned in the introduction, the communication will be used by default by SSL encryption. If the company is unable attach a direct public IP address to the server, the client provides a unique opportunity to enter a port number. In this way communication can be made compatible with VPN[10]. After the port is set on the client side the corporate network administrator appoints the selected port on the server and creates an SSL tunnel for the corporate server which is only accessible in a specific VPN network[11][12][13]. In addition, each query sent to the server is preceded by an authentication method between the client and the corporate server.

9. TESTS

It is assumed that the security can be either very strong or weak, according to the client side implementation. Because of this fact, in the test phase the security part of the product was not tested. As mentioned earlier, the enterprise server side is the weak point of this system. It offsets the flexibility and because of this the enterprise server must be able to withstand the load. The following diagram shows the test results with different size of the data. The enterprise server runs on a Tomcat 7 server? with the default configuration and the processed payload stored in MySQL. The virtual server configuration used is AMD Phenom(tm) II X6 1090T Processor and 512MB of RAM. The network bandwidth between the mobile client and the enterprise server is roughly the same and it does not affect the processing time, so the results were not compensated for the current transmission rate.

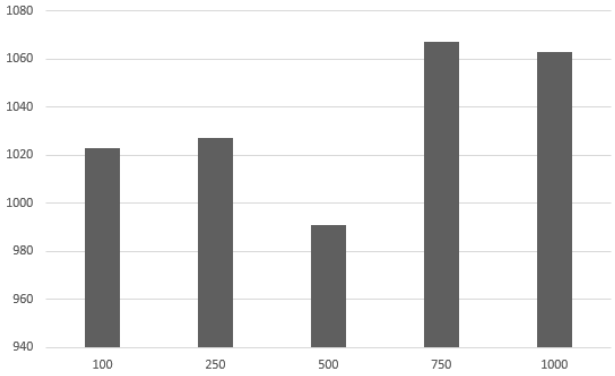


Figure 2. Response time in milliseconds according to the processed records

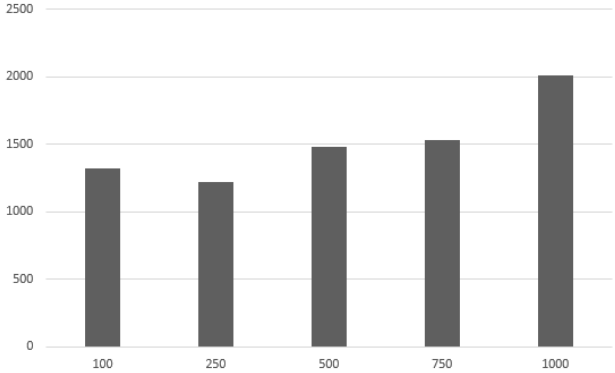


Figure 3. Response time in milliseconds according to the processed records with JOIN

The diagram (Figure 2) shows the processing time of a single SELECT SQL instruction without JOIN. The diagram (Figure 3) shows the processing time of a single SELECT SQL instruction with JOIN. The authors tried to model the general use of the product so query of sales

data in was made two different countries joined by ID. The enterprise server processes the requests (in this case the servers are on the same physical machine) and after the join it stores the result in the temporary local table. Because of bandwidth issues and the client side screen size the response contains only (a maximum of) 700 diagram points. The mobile app shows a large diagram from the response data and a small one which has two slide bars specifying which part one wants to see on the large chart. After setting an arbitrary part the client makes a query for the server to send the accurate points of the selected section. The following image shows the response time after the section is selected. The difference in response time may be caused by the network accuracy, in any case the processing time on the server side is roughly the same in case of the section selection where the result table is already in the server's temporary table.

.

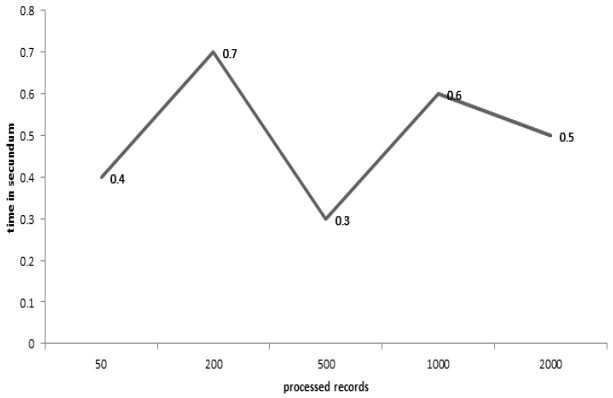


Figure 4. Chart section selection update time

SUMMARY

In this paper the authors described a design that uses a new approach for reporting systems. This approach brings a large change from the ordinary. A clearly good change is the flexibility. The new design applies a flexible structure that accommodates the specific needs of the users. It allows versatility, which means that it is not only applicable in the corporate sector. For example, real time statistics in the sport industry is more and more important these days. The system design is invented to be as safe as possible. If the service apply is correct the security is better than the existing solutions, because the systems practical safety depends on the end-user security requirements and application. The distributed design is “power save” for the service provider because it could provide a full range of services even if in the case of little start-up capital. There is no need for large hardware investments to compute a large amount of data and satisfy the users.

In addition to the benefits there are also disadvantages. Different usage needs different resource utilization which makes the corporate server side resource management very important, and the users are often not aware of the requirements. They are responsible for their own security, therefore comprehensive and thorough examination must be carried out before the use of the service. Each user should consider the possibilities and choose the product depending on that. Due to the system flexibility and integrated security the product based on this design can be a useful for the corporate users and it is definitely a good choice for the service providers because they do not need any significant resources to allocate for every client transaction. Obengo (what is available at www.obengo.com) is a new product that uses this design approach.

LITERATURE

[1] FI-WARE_Architecture. December 12, 2015. https://forge.fiware.org/plugins/mediawiki/wiki/fiware/index.php/FI-WARE_Architecture

[2] Future Internet Socio Economics Working Group - FISE Position Paper - David Hausheer, Pekka Nikander, Vincenzo Fogliati -2009

[3] Future Internet "Shaping Policies for a Digital World: The Seoul Declaration for the Future of the Internet Economy". OECD. 2008. Retrieved October 15, 2011.

[4] "Clean Slate Design for the Internet". Interdisciplinary research program website. Stanford University. Retrieved October 15, 2011

[5] David Orenstein (March 14, 2007). "A broad-based team of Stanford researchers aims to overhaul the Internet". Stanford report. Retrieved October 15, 2011.

[6] 1st IFIP/IEEE International Workshop on Management of the Future Internet (ManFI 2009),

[7] Specifications for a Secure Hash Standard (SHS) – Draft for proposed SHS (SHA-0)

[8] RFC 6234: US Secure Hash Algorithms SHA and SHA-based HMAC and HKDF. Contains sample C implementation.

[9] FIPS 180-4: Secure Hash Standard (SHS)– Current version of the Secure Hash Standard (SHA-1, SHA-224, SHA-256, SHA-384, and SHA-512), March 2012

[10] IP Based Virtual Private Networks, RFC 2341, A. Valencia et al., May 1998

[11] Point-to-Point Tunneling Protocol (PPTP), RFC 2637, K. Hamzeh et al., July 1999

[12] Phifer, Lisa. "Mobile VPN: Closing the Gap", SearchMobileComputing.com, July 16, 2006.

[13] RFC 2917, A Core MPLS IP VPN Architecture

Measurement QoS Parameters of VoIP Codecs as a Function of the Network Traffic Level

Jugoslav Jocić*, Zoran Veličković**
* Telekom Srbija AD, Prokuplje, Serbia
** College of Applied Technical Sciences, Niš, Serbia
jocic.jugoslav@gmail.com, zoran.velickovic@vtsnis.edu.rs

Abstract—**This paper analyzes the effect of the level of network traffic to the implementation QoS parameters of VoIP service in the LAN. We analyzed the number of dropped packets and packet delay variation for ten simultaneous VoIP connections. Manage Connections are made using the Asterisk VoIP server and several typical codecs are analyzed. Basic QoS parameters of VoIP services are measured in terms of variable network traffic. Measuring results clearly indicate that the number of dropped packets increases with increasing intensity of network traffic, and that very little depends on the user's VoIP codecs. On the other hand, the variation of packet delay crucial role has a selection of VoIP codecs is shown. In the realized experiment, all QoS parameters were in the preserved required limits.**

I. INTRODUCTION

The rapid expansion and distribution of computer networks has made them very suitable medium for the transmission of different kinds of content. In addition to the exchange of documents in text format, computer networks have become the main medium for the transmission of all types of multimedia content. Computer networks are not designed for multimedia content in real time, but modern communication protocols ensures the implementation of appropriate application's QoS (Quality of Service) [1]. This fact has contributed to that the exchange of digital multimedia content becomes the dominant form of network traffic. The LAN networks that are based on TCP/IP protocols cannot guarantee QoS, but they provide only "best effort" QoS services. Thus, IP networks can provide to the multimedia applications only those resources that are at that moment available. It is evident that in the multi-services networks is not easy to achieve the required application QoS. Since different applications require different QoS, the distribution of available network resources is a major research challenge [2], [3]. For voice transmission over computer networks was developed VoIP (Voice over Internet Protocol) technology that is based on the communication protocol IP (Internet Protocol). Unlike traditional telephone services, which can be realized, by circuit-switched, VoIP service is implemented with packet switching. Each IP packet among other things in the header contains the source and destination IP address, while the rest of the pack carries the application data. Network routes for delivering the packets to the destination depend on many factors, and in this paper, we analyzed the dependence of the type and intensity of network traffic. Different routes packets through the computer network as a result of the routing asymmetry can cause side effects and the VoIP service [4]. Thus, in IP networks that can happen: package does not reach the destination, that packet arrives too late, that the package arrives damaged, you arrive at your destination two identical packets, or packets arrive at their destination in the order in which they were sent.

This paper analyzes effect of the network traffic intensity in LAN to the VoIP service performance. Unlike data transfer, voice service is real time application and sets stringent QoS requirements in connection with delay and delay variation (jitter) package. On the other hand, speech shows greater tolerance to packet loss in relation to the data transfer. Providing appropriate QoS for multimedia applications is the basic problem which to be solved in VoIP networks. The implementation of appropriate QoS in VoIP networks is extremely important because in this way provides a normal conversation participants. For the implementation of VoIP services in digital networks, the most important parameters are: a) packet loss, b) packet delay and c) packet delay variation [2]. In this paper we do not discuss methods for the implementation of QoS (they are embedded in the analyzed protocols), but the performances of VoIP services based on measurable technical parameters such as packet delay variation and packet delay estimated. For the provision of high quality services, computer network must achieve the aforementioned QoS parameters within the limits defined class of service. Acceptable delay packet of speech is 150 ms, while for international connections this parameter can be tolerated in the range of 150 ms to 400 ms. The values of packet delay of 400 ms are unacceptable. Packet loss up to 1% and the jitter value of 30 ms is acceptable [1], [2], [3], [4]. The paper also analyzed the impact of VoIP codecs used in consideration of QoS parameters. The effectiveness of the applied compression algorithm is very importance because it directly affects the packet delay.

Besides the objective QoS parameters that are used in this paper, can be used subjective methods of evaluating QoS. Subjective assessment of quality speech MOS (Mean Opinion Score) is implemented in a controlled environment with the participation of a large number of listeners [5], [6]. PESQ (Perceptual Evaluation of Speech Quality) is an objective measure of the quality of the speech signal based on a comparison of the signal at the entrance and exit of the VoIP system [7]. PESQ actually

TABLE I.
BASIC FEATURES OF VOIP CODECS

VoIP Codecs	Sampling (kHz)	Bandwith (kb/s)	RAM (ms)	CPU load (MIPS)	PESQ
G.711	8	64	10	0.5	4.3
G.722	16	48,56,64	10	14	4.1
G.729	8	8	10	22	3.8
GSM	8	13	20	5	3.4
iLBC	8	15.2, 13.3	30	15,18	3.8

gives a good objective assessment of the MOS estimates but requires specialized equipment. The structure of the paper is as follows. The second chapter provides an overview of the characteristics of the considered VoIP codecs. In the third chapter describes the network protocols that transport voice signals. Networktopology and objective way of measuring QoS parameters are presented in the fourth section, while the fifth section presents the results of measurements. The final section presents conclusions based on the obtained results.

II. BASIC FEATURES OF VOIP CODECS

This paper analyzes the following VoIP codecs: G.711, G.722, G.729, iLBC and GSM. Selected codecs are very different by coding techniques, the required minimum bandwidths and processor loads. Selected codecs on the server side (VoIP server) and client-side (softphone) VoIP applications are supported. Table I shows the basic characteristics of the analyzed VoIP codecs. G.711 codec is used in standard VoIP telephony for narrowband speech signal from 0.3 kHz to 3.4 kHz. This codec requires a large bandwidth of 64 Kbit/s and achieves a high quality speech signal from PESQ = 4.3. G.722 belongs to a group of broadband voice codecs on the frequency range of 50 Hz to 7 kHz. This codec is characterized by clarity and quality of the speech signal. The bandwidth of this codec is equal to or below that of the G.711 codec. G.729 codec achieves extremely high compression, which results in a very small bandwidth with tolerance to errors. This codec is designed to transmit narrowband speech signal and speech quality is very good. GSM codec is used in mobile telephony, but it can be used in VoIP in a low bandwidth required. The main disadvantage of this codec is relatively low quality of speech. Like GSM, iLBC codec is designed for narrowband voice range. iLBC codec is an open source solution, and requires little bandwidth with very good quality contracted through speech.

III. PROTOCOLS FOR VOIP PACKETS TRANSMISSION

VoIP applications use IP, UDP and RTP protocols for packet transmission. Considering the purpose of developing, in IP protocol is a not incorporated mechanism to control data flow, as well as procedure for the correction of received packets. In addition, the IP protocol has no mechanisms for retransmission, so that in case of transmission errors due to congestion or disconnection package may be lost. These are the reasons because the IP protocol is considering unreliable. The primary function of IP is to enable the routing of traffic through the network or to provide the best possible

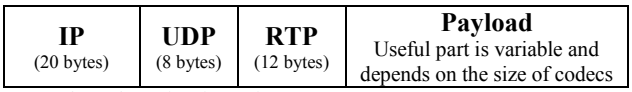


Figure 1. The structure of VoIP packets.

path from the source to the destination, using a single IP address of the participant concerned. Besides the already mentioned, a protocol enables IP fragmentation data that is necessary for various network specifics in terms of the length of the package [8]. UDP is a transport layer protocol suitable for real-time communication, so it is applied in applications for voice transmission. UDP is a connectionless protocol, which increases its efficiency due to the absence stages of establishing and closing the connection. However, UDP does not provide a mechanism for confirming delivery of data to the destination, as well as maintaining correct order of packets. Because of the displayed weakness, UDP in conjunction with RTP is often used. RTP solves problems for that UDP was not designed. RTP protocol provides the transfer function of end-to-end network necessary for the implementation of QoS for multimedia applications. Applications on the receiving side can detect packet loss, jitter or packet delay based on information provided by RTP. Each VoIP package consists of two components: a) headers and b) speech samples - the usable portion - payload. The structure of a VoIP packet is shown in Fig. 1. VoIP packet header is constant length of 40 bytes [9], whiles the useful part of the variable length and depends on the used codec. With increasing length packets, increases the efficiency of transmission, but reduces the bandwidth for other VoIP connection, which will result in packet delay. Calculating the optimal length of VoIP packets ($VoIP_{ps}$) can be determined by (1) [10]:

$$VoIP_{ps} = \left(f_s * bps * \frac{p_t}{8} \right) + S_H \quad (1)$$

where f_s is sampling frequency, *bps* indicates the number of bits per sample, p_t time packetization, and *SH* size of the header. On further increase VoIP packets can affect the type of media used and the use of security or tunnel protocol that adds its header information [11].

IV. TOPOLOGY SYSTEM AND METHOD OF MEASUREMENT

A. Computer Network

A simplified computer network topology in which the experiments were carried out is shown in Fig. 2. The foundation of presented computer network is LAN network in the *High Technical School of Applied Studies Niš* (VTŠ Niš), which owns a number of routers, switches and user computers. VoIP network consist of the VoIP server, twenty computers equipped with VoIP accessories and a separate PC for generating the network traffic. All devices are equipped with Fast Ethernet 100Mb/s network adapters and connection unleashed UTP cables. The network has the Asterisk VoIP server. The client part of VoIP communication is implemented with *CounterPath Eyebeam* softphone. Generating additional

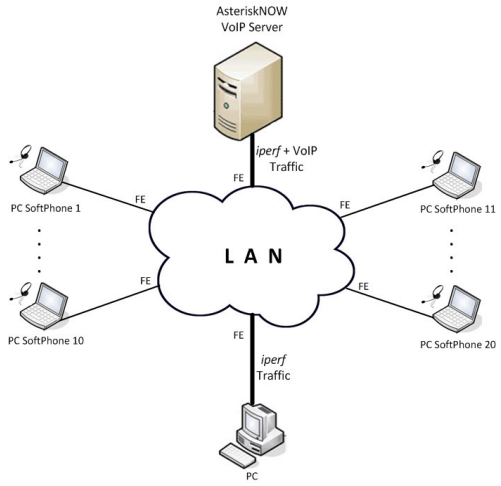


Figure 2. Block diagram of measuring topology.

traffic is carried by *iperf* program, while the measurement of QoS parameters is carried Wireshark program.

B. Asterisk VoIP server

As a VoIP server is used Asterisk server with specialized software *AsteriskNOW* that is installed on anon a standard PC. *AsteriskNOW* software is open source solution developed for the Linux operating system. Access to the Asterisk server is provided through the Web interface *FreePBX* and it was done by adjusting the *SIP* (Session Initiation Protocol) accounts as well as the activation of the codecs to be used in the experiment. To start, maintaining, and terminating sessions are also used SIP protocol [12].

C. CounterPath eyeBeam

At 20 computers that are running Microsoft Windows XP, as well as VoIP client is installed softphone CounterPath Eyebeam. SIP accounts of this program are set by the configuration of network parameters and SIP accounts AsteriskNOW server. In the present experiment was achieved ten simultaneous VoIP connection, so that all VoIP packets serve AsteriskNOW server [13].

D. Iperf

As already mentioned, to generate additional network traffic intensity in the VoIP network used is a software-tool called *iperf*. *Iperf* is also a tool for measuring network performance that is designed to run on various platforms such as Windows, Linux, UNIX or Android. *Iperf* program is installed on two computers designated as *AsteriskNOW* and *PC* (Fig. 2). On both computers are running the client and server application component, so that both computer generated and accept UDP traffic [14]. In order to achieve the desired level of network traffic should set the appropriate parameters via *iperf* command on the client as follows:

iperf -c -u -b 90m 192.168.1.10 -t 600 s 1,

while on the server side should set the appropriate parameters iperf command:

iperf -s -u –i,

first -c Parameter indicates the part of the client, while -s indicates the server part. IP address 192.168.1.10 is the address of the computer that accepts generated traffic, -u indicates the type of generated packets - UDP. Parameter -b 90m provides generating traffic from 90Mb/s, while the parameter -t 600 defines the total time interval of sending packets of 600s. Display interval is defined -i parameter in this case is 1s.

E. Wireshark

Wireshark is open source software and belongs to a class of network protocol analyzer. The main function of this software tool is to record the packets on a network interface as defined and view the captured packets for analysis [15]. For purposes of measuring QoS parameters in the local network, such as packet loss and jitter, using a software package Wireshark. Because the variation of packet delay only measured at the receiver side, it should be noted that Wireshark collects data based on information from the RTP packet. Wireshark is installed and running before the establishment of VoIP connections on all computers marked with the PC SoftPhone 11 to PC SoftPhone 20.

V. Results of Measurement

A. Measurement Procedure

An experiment in which the measured QoS consisted of 10 simultaneous VoIP connections in a time of 10 min. We implemented a series of seven measurements. First measurement was performed without additional network traffic, while the other six measurements are performed with additional network traffic. Software tool iperf generates additional network traffic from 20Mb/s, 50Mb/s, 70Mb/s, 80Mb/s, 90Mb/s and 95Mb/s between Asterisk VoIP server and PC. Generating an additional network traffic between iprf and VoIP servers is simulated desired level of network traffic. While this is not the typical VoIP situation in the LAN, this is a way to obtain a valid measurement data. The program Wireshark analyses the packet losses (lost or discarded) and jitter of all VoIP connections. At eight VoIP simultaneous connections are applied only voice codecs, while the other two besides VoIP codecs and video codecs implemented H.263 and H.264. We used the already mentioned VoIP codecs: G.711, G.722, G.729, GSM and iLBC.

B. Measured data

In the performed experiments, the number of dropped packets is measured, as well as average and maximum jitter value of all VoIP connections. In Fig. 3 shows the number of dropped packets, in Fig. 4 the average value, and Fig. 5 maximum jitter of VoIP packets. We analyzed the following VoIP codecs G.711, G.722, G.729, GSM, iLBC with variation of network load. Packet loss represents packets that have not arrived at their destination, and derive from errors in the network, i.e., damaged packages that usually occur because of LAN overloading. Fig. 3 shows the percentage of lost packets, which are estimated in relation to the total number of sent

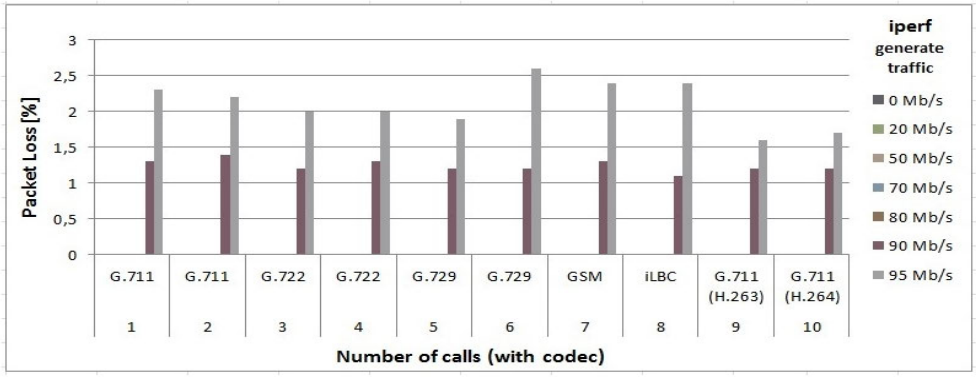


Figure 3. Packet loss for codecs G.711, G.722, G.729, GSM, iLBC

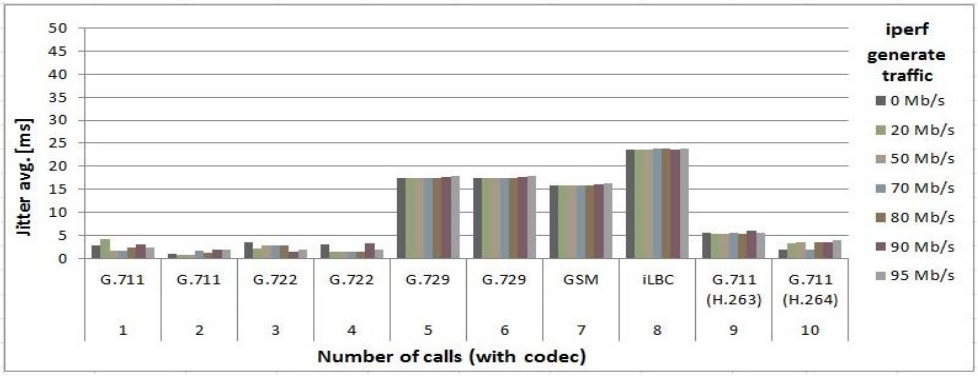


Figure 4. Average time jitter for codecs G.711, G.722, G.729, GSM, iLBC

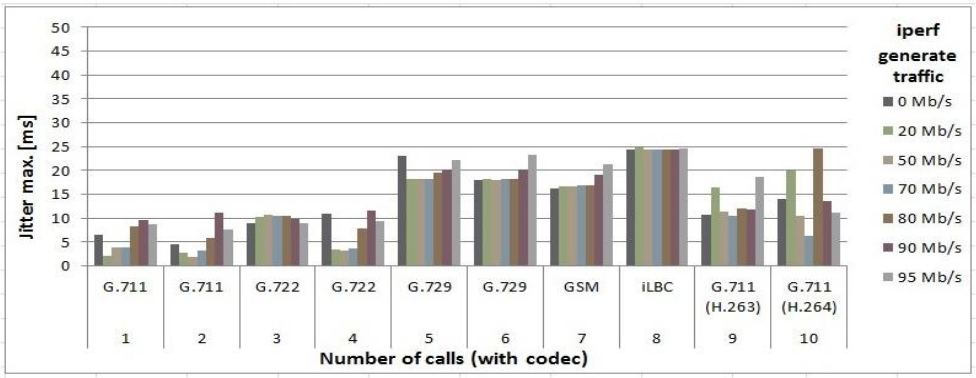


Figure 5. The maximum time jitter for codecs G.711, G.722, G.729, GSM, iLBC

packets. From the graph it can be clearly established that packet loss is not detected when the network load is less than 80Mb/s.

The loss of packets of all codecs can be observed when the network traffic is greater than 90Mb/s. The percentage of packet loss is in the range from 1.1% to 1.4%, while the network load of 95 Mb/s packet loss percentages was in the range from 1.6% to 2.6%. At this level of network traffic at all codecs, the percentage of packets dropped considerably increased. From Figure 3 it can be observed that for some VoIP connections percentage of packets dropped nearly doubled (connection no. 6).

Jitter is a variation of packet delay, i.e., the difference in time delay packets in the same stream. Average time jitter for codecs G.711 and G.722 when the experiment used video codec, i.e., without it, there was no greater than 6ms (Fig. 4). Significantly higher jitter is observed for codec G.729 (8ms), while for the GSM codec jitter was 16ms. The average value of jitter for iLBC codec was 24ms, which is the worst result. It is important to note that in all the codecs maximum jitter value did not exceed 25 ms can be seen from Fig. 4.

Based on these results, we can conclude that the average and maximum jitter values for all codecs are not greater than 30 ms, which means that they are within recommended limits.

However, it is evident that jitter values depend on the used VoIP codecs. Percentage of packet loss values are in the recommended values for the load in the network [16] and VoIP servers up to 80Mb/s. Packet loss is greater than the recommended 1% were recorded for loads over 90Mb/s. From the graph, it can be concluded that packet loss very little depends of the selected VoIP codecs.

VI. CONCLUSION

The paper presents the impact of the level of network load into a local network on technical QoS parameters of voice codecs. Based on experiments it was found that very little packet loss depends on the type of applied voice codec. Significant impact on packet loss had a level of network traffic. Measured data indicate that the packet loss occurred only when the level of network traffic approaching the designed capacity of the connection. Packet loss measurements at 6 and 7 are higher than 1%, which had a negative impact on the quality of the received speech. By measuring was determined that jitter does not depend on the network load and the VoIP server, it depends only of the type of applied voice codec. The values of the measured jitter were less than 25ms and it had no significant effect on the quality of the received speech.

The measurements from 1 to 5 have QoS parameters for VoIP codecs within recommended limits. In this paper, it is shown that it is possible to realize quality VoIP service in LAN using standard VoIP codecs: G.711, G.722, G.729, GSM and iLBC. However, when the level of network traffic approaches to the projected capacity, it may cause increased packet loss and jitter, which can have the effect of variable quality VoIP service. The experiment shows that in this limit loads jitter remains within specified limits. Given that the packet loss and jitter occur only when the network load greater than 80 Mb / s, it can be concluded that the implementation of VoIP service is not a problem in this way conceived LAN.

REFERENCES

[1] T. Szigeti,C. Hattingh, "*End-to-End QoS Network Design: Quality of Service in LANs, WANs, and VPNs*", Cisco Press, 2004.

[2] End-user multimedia QoS categories, *ITU-T Recommendation G. 1010*, 2001.

[3] M. Jevtović, Z. Veličković, Komunikacioni protokoli prepletenih slojeva, Akademska misao, 2013.

[4] International Telecommunication Union, "*ITU-T Recommendation G.114, One-way transmission time*", 2004.

[5] Z. Veličković, Z. Milivojević, "*MOS test baziran na Webu*", YUINFO 09, Ref. 025.pdf, ISBN 978-86-85525-04-9, Kopaonik, 2009.

[6] International Telecommunication Union, "*ITU-T P.800, - Methods for Subjective Determination of Transmission Quality*", 1996.

[7] International Telecommunication Union, "*ITU-T P.862, Perceptual evaluation of speech quality PESQ*", 2001.

[8] B. Goode, "*Voice over Internet Protocol*", Proceedings of the IEEE, 2002.

[9] Cisco Press, "*Quality of Service for Voice over IP*", 2001.

[10] Z. Bojović, Z. Perić, V. Delić, E. Šećerov, M. Sečujski, V. Šenk, "*Comparative Analysis of the Performance of Different Codecs in a Live VoIP Network using SIP Protocol*", ELECTRONICS AND ELECTRICAL ENGINEERING No. 1(117), 2012.

[11] Z. Veličković, M. Jevtović, V. Pavlović, "*Quality of services in IP/MPLS networks*", UNITECH 2013, pp. II-107-112, Gabrovo 2013.

[12] C. Hattingh, D. Sladden, Z. Swapan, "*SIP Trunking*", Cisco Press, 2010.

[13] "*eyeBeam 1.5 for Windows User Guide*", CounterPath Corporation, 2007.

[14] Iperf - The TCP_UDP Bandwidth Measurement Tool, *https://iperf.fr*, 2014.

[15] U. Lamping, R. Sharpe, E. Warnicke, "*Wireshark User's Guide* (*For Wireshark 1.99*)", 2014.

[16] M. Jeftović, Z. Veličković, "*Kvalitet usluga digitalnih mreža*", Akademska misao, 2014.

An Efficient MATLAB Implementation of OFDM/OQAM Modulator with Orthogonal Pulse Shaping Filters

Selena Vukotić, Desimir Vučić
Faculty of computer science, Belgrade, Serbia
Faculty of computer science, Belgrade, Serbia
svukotic@raf.edu.rs, dvucic@raf.edu.rs

***Abstract*— In this paper we made a theoretical overview considering OFDM/OQAM signals in continuous and discrete time and we presented an orthogonality conditions for pulse shaping filters in discrete time. By introducing the efficient scheme of the OFDM/OQAM modulator [1], we gave our implementations of OFDM/OQAM modulator in MATLAB respecting the orthogonality conditions [3]. Based on the simulations results, power spectrums for different number of subcarriers and pulse shaping filters are shown.**

I. INTRODUCTION

The Orthogonal frequency division multiplexing (OFDM) signal, as one of multi-carrier modulation (MCM) signals, has very important place in modern telecommunications because of many good features and advantages over single-carrier signals. It is a part of numerous standards: European digital audio broadcasting (DAB), digital video broadcasting (DVB) standards, asymmetric digital subscriber line (ADSL), WLAN (IEEE 802.11 a/g), WiMAX (IEEE 802.16), 3G LTE, 4G wireless standards and others.

OFDM systems offer robustness in the situations where multipath propagation and narrowband interference occur, which is a quite common scenario in a radio environment. Also, very good property of those systems is high spectral efficiency, which could be maximal in a case of critical sampling. An important task considering OFDM systems is certainly the choice of pulse shaping filters, because good time-frequency localization of those filters increases performance of the system in time-dispersive or frequency-dispersive channels.

Maximal spectral efficiency could be achieved in the case of critical sampling for both OFDM systems based on ordinary quadrature amplitude modulation (OFDM/QAM) and OFDM systems based on offset quadrature amplitude modulation (OFDM/OQAM). However, OFDM/OQAM offers wide opportunities in a choice of pulse shaping filter and, thereby, better time-frequency localization of pulse shaping filter. As the matter of fact, it has been shown that OFDM/QAM system with good time-frequency localization (i.e. pulse shaping filter meets the condition of compactly support function) and with orthogonal subcarriers could not achieve maximal spectral efficiency [4].

On the other side, it was shown [1, 2, 3] that OFDM/OQAM systems can meet discrete-time orthogonality conditions preserving maximal spectral efficiency and they could also have good time-frequency localization [4].

In this paper we put together all three previously mentioned properties of OFDM/OQAM and we presented an implementation of the computationally efficient scheme of OFDM/OQAM modulator [1], where pulse shaping filters are well-localized and they meet discrete-time orthogonality conditions [3] and spectral efficiency is maximal. The orthogonal filters were made by discrete-time Zak transform (DZT), which is also very computational efficient procedure. Our contribution is an efficient MATLAB implementation of OFDM/OQAM modulator, which combines good computational properties of two methods, described in [1] and [3].

II. THEORETICAL BASIS

A. OQAM/OFDM Signal in Continuous and Discrete Time

The continuous-time baseband OFDM/OQAM signal, with $2M$ subcarriers, could be written as follows [1]:

$$s(t) = \sum_{n=-\infty}^{+\infty} \sum_{m=0}^{M-1} \left(c_{2m,n}^{R} p(t - nT_0) + jc_{2m,n}^{I} p\left(t - \frac{T_0}{2} - nT_0\right)\right) e^{j2\pi(2m)F_0 t} + \left(jc_{2m+1,n}^{I} p(t - nT_0) + c_{2m+1,n}^{R} p\left(t - \frac{T_0}{2} - nT_0\right)\right) e^{j2\pi(2m+1)F_0 t}. \quad (1)$$

In (1) T_0 represents the signaling interval, $F_0 = 1/T_0$ is the spacing between two successive carriers, $c^{R}_{m,n}$ and $c^{I}_{m,n}$ are the real and imaginary parts, respectively, of the QAM complex-valued symbols and $p(t)$ is a pulse shape filter impulse response.

In order to get simplified expression, the following notation could be used:

$$a_{2m,2n} = c_{2m,n}^{R}, \quad a_{2m,2n+1} = c_{2m,n}^{I}, \quad a_{2m+1,2n} = c_{2m+1,n}^{I}, \quad a_{2m+1,2n+1} = c_{2m+1,n}^{R}. \quad (2)$$

$$\varphi_{2m,2n} = 0, \qquad \varphi_{2m,2n+1} = \frac{\pi}{2},$$
$$\varphi_{2m+1,2n} = \frac{\pi}{2}, \quad \varphi_{2m+1,2n+1} = 0. \tag{3}$$

In order to get discrete-time formulation of the OFDM/OQAM signal, sampling period $T_s=T_0/2M$ was applied, which is the critical sampling period because $2M$ QAM complex valued symbols would be transmitted during symbol period T_0. So, the maximal spectral efficiency was achieved this way. In our implementation we trunked a pulse shaping filter, $p(t)$, to the interval $[-LT_s/2, -LT_s/2]$ and delayed by $(L\text{-}1)T_s/2$, which gave us a causal discrete time prototype filter. Finally, discrete time OFDM/OQAM signal can be written in the following form:

$$s(k) = \sum_{m=0}^{2M-1} \sum_{n=-\infty}^{+\infty} (a_{m,n} p(k - nM) e^{j\pi \frac{m}{M}\left(k - \frac{L-1}{2}\right)} e^{j\varphi_{m,n}}). \tag{4}$$

In order to get to an efficient implementation scheme, the variable $\varphi_{m,n}$ from (3) has been redefined:

$$\varphi_{m,n} = \frac{\pi}{2}(n + m) - \pi mn. \tag{5}$$

Also, new variable was introduced:

$$x_m^0(n) = a_{m,n} e^{j\pi n/2}. \tag{6}$$

The modification of $\varphi_{m,n}$ does not affect signal statistic. The efficient implementation scheme [1] for signal defined by (4), is presented in Fig. 1.

We need to introduce the following notation: $x_m(n) = a_{m,n}$, Z transform of $x_m(n)$ is $X_m(z)$ and $X_m(-jz)$ is the Z transform of $x^0_m(n)$. Also, $G_l(z)$ is the poly-phase component of a pulse shaping filter, given by:

$$G_l(z) = \sum_n p(l + 2nM) z^{-n}. \tag{7}$$

In [3] it was shown that pulse shaping filters, which meet discrete-time orthogonality conditions, could be obtained by DZT. This method is applicable on the signal defined by the following expression:

$$s(k) = \sum_{m=0}^{2M-1} \sum_{n=-\infty}^{+\infty} (c^R_{m,n} p(k - 2nM) e^{j\pi \frac{m}{M}\left(k - \frac{\alpha}{2}\right)} + jc^I_{m,n} p\left(k + \frac{M}{2} - 2nM\right) e^{j\pi \frac{m}{M}\left(k - \frac{\alpha}{2}\right)}). \tag{8}$$

In [3] we could find that α is a parameter of pulse shaping filter, which satisfies $\alpha \in [0, 2M\text{-}1]$ and $\alpha = (L+M\text{-}1) \bmod (2M)$ (mod stands for the modulo operation), where L is the length of the prototype filter. We shall also introduce the parameter $r \in Z$, which together with α allows

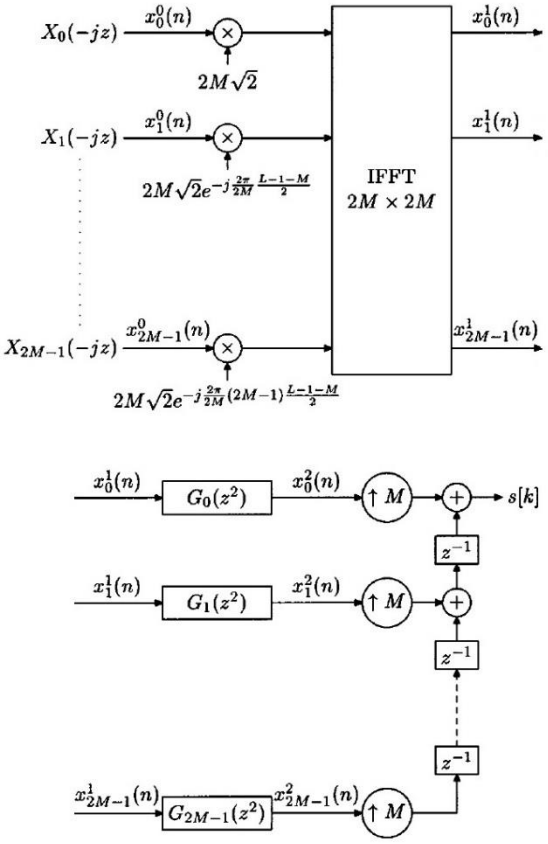


Fig. 1. OFDM/OQAM modulator realized with an IFFT

a flexible choice of the center of symmetry of the pulse shaping filter $p(n)$:

$$p(k) = p(\alpha + \frac{(2r+1)M}{2} - k). \tag{9}$$

We shall point out that $p(n)$ is necessarily the symmetrical and causal impulse response for the implementation scheme in Fig. 1, as well as in (9) in order to make orthogonal basis in discrete domain by DZT, i.e. by the method described in [3].

If we put $\alpha = (L+M\text{-}1)$ instead of $\alpha = (L+M\text{-}1) \bmod (2M)$ in (8), we will change only the sign of some $c^R_{m,n}$ and $c^I_{m,n}$ symbols according to their m index, which will not change the signal statistics. So, after replacement of $\alpha = (L+M\text{-}1)$, we got a statistically equivalent signal:

$$s(k) = \sum_{m=0}^{2M-1} \sum_{n=-\infty}^{+\infty} (c^R_{m,n} p(k - 2nM) e^{j\pi \frac{m}{M}\left(k - \frac{L-1}{2}\right)} + jc^I_{m,n} p\left(k + \frac{M}{2} - 2nM\right) e^{j\pi \frac{m}{M}\left(k - \frac{L-1}{2}\right)}) e^{j\pi \frac{m}{2}}. \tag{10}$$

If we compare the definitions of $s(k)$ by (4) and $\varphi_{m,n}$ by (5) with the definition of $s(k)$ by (10), we could conclude that it is the same signal from the statistical point of view. So, the signal defined by (10) could be generated via scheme from Fig.1 and could be processed via method described in [3] in order to achieve orthogonal basis in

discrete-time.

B. Orthogonality Conditions

The pulse shaping filter *p(k)* is sad to be orthogonal if it satisfies perfect reconstruction in the absence of a channel, i.e. $c^R_{m,n}' = c^R_{m,n}$ and $c^I_{m,n}' = c^I_{m,n}$, where $c^R_{m,n}'$ and $c^I_{m,n}'$ are detected real and imaginary parts of a symbol. We could consider the equivalent path from the (m+u)-th transmitter sub-channel to the m-th receiver sub-channel, so it follows [3] that *p(k)* is orthogonal if the following conditions are satisfied for $u \in [0, 2M\text{-}1]$, $n \in Z$:

$$s(k) = [Re\left\{p(k-n2M)e^{j2\pi\frac{u}{2M}\left(k-\frac{\alpha}{2}\right)}\right\} * p'(k)]_{k=0} = \delta(n)\delta(u), \quad (11)$$

$$s(k) = [Re\left\{jp(k+M-n2M)e^{j2\pi\frac{u}{2M}\left(k-\frac{\alpha}{2}\right)}\right\} * p'(k)]_{k=0} = 0, \quad (12)$$

$$s(k) = [Im\{p(k-n2M)e^{j2\pi\frac{u}{2M}\left(k-\frac{\alpha}{2}\right)}\} * p'(k-M)_{k=0} = 0, \quad (13)$$

$$s(k) = [Im\{\, jp(k+M-n2M)e^{j2\pi\frac{u}{2M}\left(k-\frac{\alpha}{2}\right)}\} * p'(k-M)]_{k=0} = \delta(n)\delta(u). \quad (14)$$

In previous expressions $p'(k) = p(-k)$ and operator * stands for convolution operation. For $u \neq 0$ it is guaranteed that inter-channel interference (ICI) is perfectly cancelled, whereas for $u = 0$ inter-symbol interference (ISI) is perfectly cancelled. More specifically, for $u = 0$ (12) and (13) give condition for zero ISI between real and imaginary parts of QAM symbols, while (11) and (14) guarantee zero ISI for real and imaginary parts.

It could be shown [3] that the orthogonality conditions are met if the following equation is satisfied:

$$\sum_{r=-\infty}^{\infty} p(k-rM)p(k-rM-n2M) = 1/M\delta(n). \quad (15)$$

For the final expression of the orthogonality conditions we shall need DZT of the function *p(k)*, which is defined as:

$$Z_p^M(k,\theta) = \sum_{r=-\infty}^{\infty} p(k+rM)e^{-j2\pi r\theta}.. \quad (16)$$

The inverse DZT is defined as:

$$p(k) = \int_0^1 Z_p^M(k,\theta)d\theta. \quad (17)$$

In the context of this paper, DZT is a signal transformation and it leads us to the orthogonality conditions in the DZT domain [3]:

$$\left| Z_p^M(k,\theta)\right|^2 + \left| Z_p^M(k,\theta-0.5)\right|^2 = 2/M$$

for k=0,1,…,M-1. (18)

C. Orthogonalization Procedure

Our goal was to make a symmetric function satisfying (15). We could start from an arbitrary symmetric filter which will be modified to obtain a symmetrical orthogonal pulse shaping filter and the whole procedure will be done in DZT domain.

Starting from an arbitrary filter $p(k)$ satisfying (9) and with $\alpha \in [0, 2M\text{-}1]$ and $r \in Z$, an orthogonal pulse shaping filter $p_o(k)$ can be obtained as [3]:

$$Z_{p_o}^M(k,\theta) = \frac{2Z_p^M(k,\theta)}{\sqrt{2M\left|Z_p^M(k,\theta)\right|^2 + 2M\left|Z_p^M(k,\theta-0.5)\right|^2}}. \quad (19)$$

We could notice the following:

$$Z_p^M(k,\theta-1) = Z_p^M(k,\theta). \quad (20)$$

By inserting (19) into (18) and using (20), it is easy to see that $p_o(k)$ is orthogonal function, i.e.:

$$\left| Z_{p_o}^M(k,\theta)\right|^2 + \left| Z_{p_o}^M(k,\theta-0.5)\right|^2 = 2/M$$

for k=0,1,…,M-1. (21)

D. Orthogonalization Procedure in Discrete Time-Frequency Grid

In practice we will apply DZT in discrete domain, according to:

$$Z_p^M(k,S) = \sum_{r=0}^{S-1} p(k+rM)e^{-\frac{j2\pi sr}{S}},$$

for k=0,1,…,M-1 and s=0,1,…,S-1. (22)

In (21) S is an integer constant satisfying L=MS. It obviously might be necessary to perform zero-padding on filter $p(k)$, so the filter got to the length L=MS if it was shorter at the first place. We could notice that the computation of (21) reduces to the column-wise fast Fourier transform (FFT) of the S x M matrix [3], which is presented in Fig. 2.

We could summarize the steps of orthogonalization procedure as follows:

-Design an initial filter, which meets conditions in (9);

-Apply zero-padding of $p(k)$ to obtain a filter of length L=MS;

-Compute the DZT of the orthogonal filter $p_o(k)$ according to:

$$Z_{p_o}^{(M,S)}(k,s) = \frac{2Z_p^{(M,S)}(k,s)}{\sqrt{2M\left|Z_p^{(M,S)}(k,s)\right|^2 + 2M\left|Z_p^{(M,S)}(k,s-S/2)\right|^2}} ;(23)$$

-Compute the inverse DZT to obtain the orthogonal pulse shaping filter $p_o(k)$, i.e. compute the inverse FFT of the columns of the matrix in Fig. 2.

The algorithm previously described does not guarantee that $p_o(k)$ is automatically well-localized in time and frequency. It was observed [3] that starting from a low-pass filter or a Gaussian function with bandwidth approximately equal to $1/(2M)$ leads to well-localized OFDM/OQAM pulse shaping filter.

It would be interesting to notice that the described algorithm of orthogonalization is computationally very cheap, because it is based on FFT and division in DZT domain.

III. SIMULATIONS IN MATLAB

A. Parameters for Simulations

In our simulations we applied the Gaussian and Kaiser windows as initial filters. Both filters are symmetrical, Gaussian filter has very good time-frequency localization [4] and Kaiser is optimal considering the compromise between the width of transition zone and minimal amplitude of side lobes.

All filters have the order of 111 and they were made by fir1 instruction in MATLAB. The cut-of frequency, which correspondents to the attenuation of 6 dB, was set to $1/2M$. For Gaussian window the standard deviation was set to 5^{-1} and 10^{-1}, and for the Kaiser window the parameter β was set to 30.

In our implementation the number of subcarriers can be set to $2M$-1, where M is the parameter satisfying $M \in \mathbb{N}$. We have chosen small number of subcarriers (3 and 7) in the figures presented below only because of graphical presentation of results.

As an input data for simulations, we applied 10.000 blocks of $2M$ complex valued QAM symbols, where each symbol has equal probability.

We assumed the sampling frequency of 1000 H_z and the carrier frequency of 250 H_z for the signal in transposed bandwidth. Also, for the purpose of visual presentation, we set the total bandwidth to approximately 200 H_z. So, when the number of subcarriers is smaller, the bandwidth around each subcarrier is wider and vice versa.

B. Simulation Results

We made several scenarios where all subcarriers (subcarrier 1, subcarrier 2, subcarrier 3 etc.) or only two successive subcarriers were occupied (subcarrier 1, subcarrier 2). In Fig. 3 and Fig. 4 we used Gaussian window with standard deviation of 10^{-1} and in Fig. 5 and Fig 6 with standard deviation of 5^{-1}. Power spectrum densities, obtained by Welch method, are presented below.

$$P = \begin{bmatrix} p(0) & p(1) & \cdots & p(M\text{-}1) \\ p(M) & p(M\text{+}1) & \cdots & p(2M\text{-}1) \\ p(2M) & p(2M\text{+}1) & \cdots & p(3M\text{-}1) \\ \vdots & \vdots & & \vdots \\ p(L\text{-}M) & p(L\text{-}M\text{+}1) & & p(L\text{-}1) \end{bmatrix}$$

Fig. 2. The matrix of pulse shaping filter coefficients

Fig. 3. The orthogonal filter obtained from Gaussian window, M=2 and all subcarriers

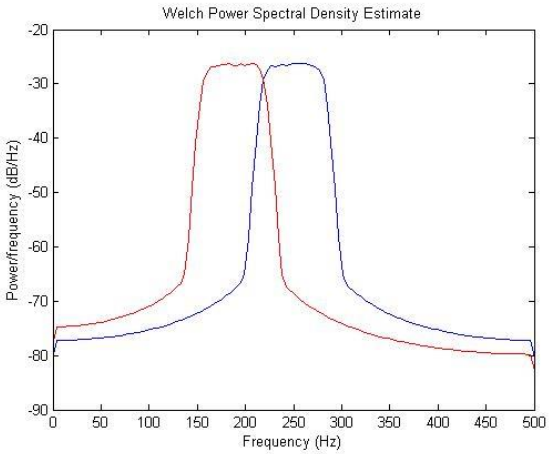


Fig. 4. The orthogonal filter obtained from Gaussian window, M=2 and subcarriers 1 and 2

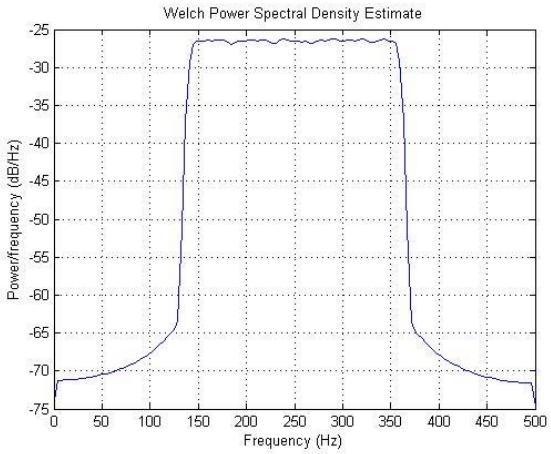


Fig. 5. The orthogonal filter obtained from Gaussian window, *M*=4 and all subcarriers

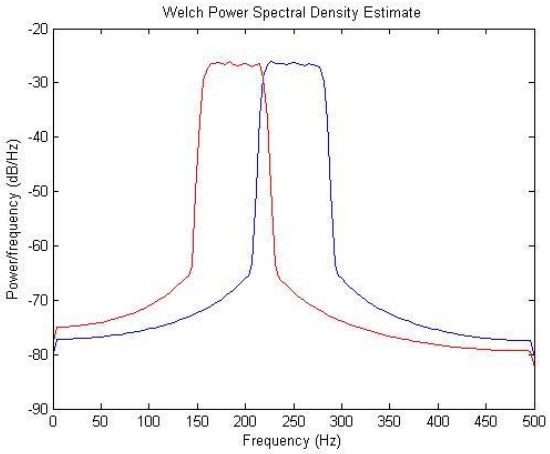


Fig. 8. The orthogonal filter obtained from Kaiser window, *M*=2 and subcarriers 1 and 2

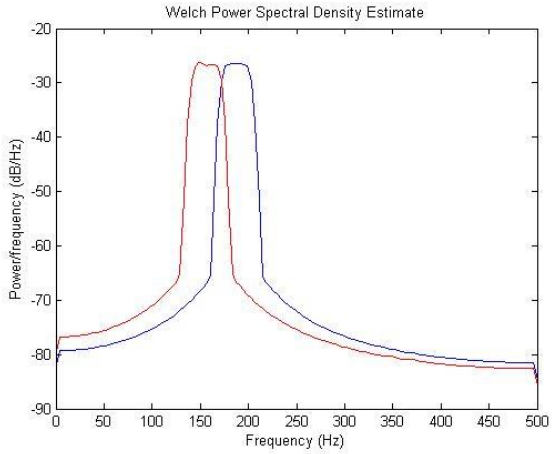


Fig. 6. The orthogonal filter obtained from Gaussian window, *M*=4 and subcarriers 1 and 2

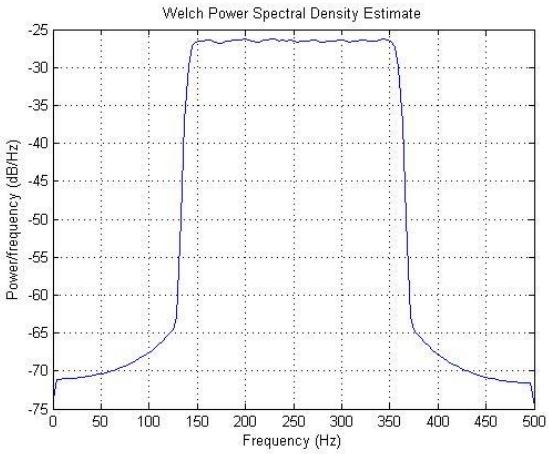


Fig. 9. The orthogonal filter obtained from Kaiser window, *M*=4 and all subcarriers

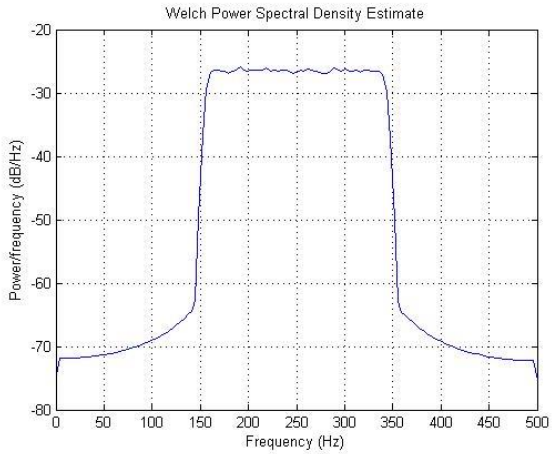


Fig. 7. The orthogonal filter obtained from Kaiser window, *M*=2 and all subcarriers

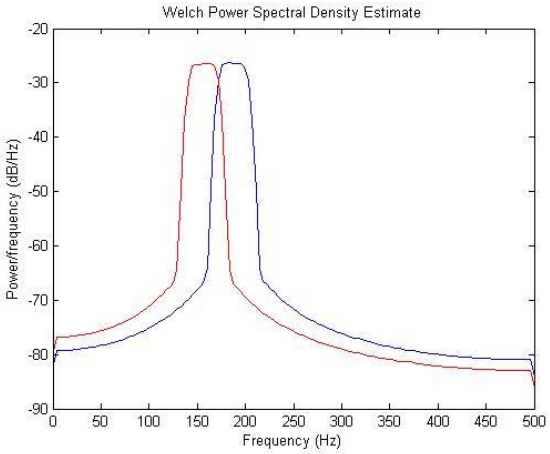


Fig. 10. The orthogonal filter obtained from Kaiser window, *M*=4 and subcarriers 1 and 2

IV. MATLAB CODE

The part of MATLAB code, which calculates the orthogonal filter for initial symmetrical filter and number of subcarriers as input arguments, is shown below. For more information on code, interested readers refer to [5].

```
function [ filteroutput ] =
orthogonalization( filterinput,M)
%M - the number of subcarriers
%filterinput - symetrical filter
L=length(filterinput);
K=ceil(L/(M/2));
if mod(K,2)==1
    K=K+1;
end
Lg=M/2*K;
if L<Lg
    filterinput(Lg)=0;
end

index=1:M/2:Lg-M/2+1;
G=filterinput(index)';
for i=2:M/2
    G=[G filterinput(index+i-1)'];
end
Zg=zeros(K,M/2);
for i=1:M/2
    Zg(:,i)=fft(G(:,i));
end
Zgpom=[Zg;Zg];
Zgt=zeros(K,M/2); %translated matrix
for i=1:K
    Zgt(i,:)=Zgpom(i+K/2,:);
end
Zgo=2*Zg./(sqrt(M*(abs(Zg)).^2+M*(abs(
Zgt)).^2));
for i=1:M/2
    go(:,i)=ifft(Zgo(:,i));
end
filteroutput=reshape(go',1,Lg);
end
```

V. CONCLUSION

This paper connects an efficient MATLAB implementation of an OFDM/OQAM modulator and also computationally efficient method of projecting orthogonal filter basis in discrete time. We have examined OFDM/OQAM signals, which allow time-frequency well-localized pulse shaping filters, even for the case of critical sampling, i.e. maximum spectral efficiency, in contrast to classical OFDM/QAM [6]. Beside that, pulse shaping filters in our simulations meet discrete-time orthogonality conditions, which preserves the signal from distortion caused by truncating of infinitely pulses. Our OFDM/OQAM modulator could be a good starting point for various examinations of this signal, which is becoming more and more popular because of some advantages over classical OFDM/QAM.

ACKNOWLEDGMENT

This work was supported by the Serbian Ministry of Education and Science under technology development project TR32028 – “Advanced Techniques for Efficient Use of Spectrum in Wireless Systems”.

REFERENCES

[1] P. Siohan, C. Siclet and N. Lacaille, “Analysis and design of OFDM/OQAM systems based on filterbank theory” IEEE Transaction on Signal Processing, vol. 50, no. 5, may 2002.

[2] H. Zhang, “Filter bank based multicarrier (FBMC) for cognitive radio systems“ (PhD dissertation), *Docteur du Conservatoire National des Arts et Métiers et Wuhan Université*, 2010.

[3] H. Bölcskei, “Orthogonal frequency division multiplexing based on offset QAM ,” *Book chapter in "Advances in Gabor Analysis"*, H. G. Feichtinger and T. Strohmer, eds., Birkhäuser, pp. 321-352, 2003.

[4] J. Du, S. Signell, “Classic OFDM systems and pulse shaping OFDM/OQAM systems,” Royal Institute of Technology SE-100 44 Stockholm, Sweden, February 2007.

[5] S. Vukotić, D. Vucic, “An efficient MATLAB implementation of OFDM/OQAM modulator”, *IcETRAN, Serbia, EKI2.3*, June 2014.

[6] H. Bölcskei, “Blind estimation of symbol timing and carrier frequency offset in wireless OFDM systems,” *IEEE Transaction on Communications*, vol. 42, NO. 6, June 2001

Smart City Services for Citizen-Centric Internet of Things

Nenad Gligoric*, Srdjan Krco*, Dejan Drajic*, Ignacio Elicegui**, Carmen López**, Luis Sánchez**, Michele Nati***, Jorge Bernal Bernabé****, José L. Hernández-Ramos****,Davide Carboni*****, Alberto Serra*****

*DunavNET, Research and Development, Antona Cehova 1, 21000 Novi Sad, Serbia

**University of Cantabria, 39005 Santander, Cantabria, Spain

***Centre for Communication Systems Research, University of Surrey, Guildford, GU2 7XH, Surrey, UK

****University of Murcia, Department of Information and Communications Engineering
Campus de Espinardo, 30100 Murcia, Spain

*****Information Society Area, CRS4
Piscina Manna, Edificio 1 - 09010 Pula (CA) – Italy

{nenad.gligoric, srdjan.krco, dejan.drajic}@dunavnet.eu, {iemaestro, clopez, lsanchez}@tlmat.unican.es, m.nati@surrey.ac.uk, {jorgebernal, jluis.hernandez}@um.es, {dcarboni,alserra}@crs4.it

***Abstract*— SocIoTal project addresses a crucial next step in the transformation of an emerging business driven Internet of Things (IoT) infrastructure into an all-inclusive one for the society by accelerating the creation of a socially aware citizen-centric IoT. In this paper are described the scenarios selected for the field trials and pilot deployment in this project, together with the evaluation methodology. The purpose of field trials and pilots is to test the developments over real environments, with real users, facing all the constraints and limitations that a complex society can pose in these kinds of trials. Within the pilot's evaluation process different methodologies and tools are considered: questionnaires and qualitative interviews from target groups collected during workshops where the services are presented to the end users, and real life testing.**

I. INTRODUCTION

SocIoTal project addresses a crucial next step in the transformation of an emerging business driven Internet of Things (IoT) infrastructure into an all-inclusive one for the society by accelerating the creation of a socially aware citizen-centric IoT. By providing adequate socially aware tools and mechanisms that simplify complexity and lower the barriers of entry it will encourage citizen participation in the IoT. In our previous work [1][2] we have highlighted the challenges that the creation of a privacy-aware framework needs to face for envisioning the social perspective of citizen-centric services based on the IoT paradigm. Later on, we have published [3] analysis and definition of use cases as a result of co-creation workshops and feedback received from the citizens. A logic step forward (which is also a main topic of this paper) is a definition of pilots and field trials going to be developed and evaluated.

In this paper we describe the services and pilots to be deployed during the last year of SocIoTal, as well as the evaluation process and the test defined for each one to evaluate its correct execution. The selected pilots come from the work done in previous phases of the project, evolved through the progress in the rest of the SocIoTal work packages and refined with several meetings and co-working sessions. As the result, but still having further to go, selected pilots will provide scenarios to play and test all the innovations introduced by SocIoTal, as well as the selected platforms to build its running instantiation. The purpose of field trials and pilots is to test the developments over real environments, with real users, facing all the constraints and limitations that a complex society can pose in these kinds of trials.

The paper is organized as follows. In Section II, SocIoTal platform is described, field trials are explained and evaluation methodology of trials is provided. Section III presents the selected SocIoTal service pilots. In Section IV KPIs and evaluation questionnaires are given. The paper is concluded with section V.

II. IoT PLATFORM

The SocIoTal platform uses the principles and the framework defined in the Internet of Things – Architecture (IoT-A) referent model [4]. This approach allows an existing well-showed module to be reused and extended with privacy-oriented features.

As one of the SocIoTal goals is to enable two main user groups (citizens and developers) to use its services, platform is composed of functional blocks enabling both to use the platform. Accordingly, an APIs (Application Programming Interfaces) that expose a set of the required functionalities of the SocIoTal platform to application developers are defined.

For citizens, the SocIoTal User Environment is envisioned as service composed of two applications: a mobile application, i.e. Mobile UserEnv; and a web based application, the Web User Environment. Both, the web application and the mobile application, are the front-end to an API-to-API broker that enables data and events to flow either from device-to-device or from device-to-service (e.g. from a sensor to a social network) having in

the user workspace the possibility to compose events and actions in a straightforward way (Figure 1).

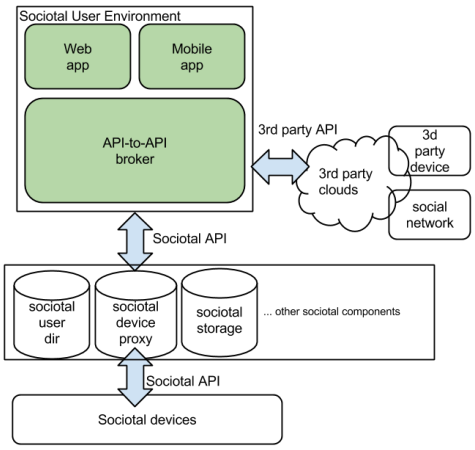


Figure 1. High level view of components and subcomponents of UserEnv in the context of SocIoTal and 3rd party systems

A. Field Trials

In this section evaluation methodology is described as well as field trials to be deployed within the project in order to test the different tools and enablers developed within SocIoTal.

1) Registering people, devices and resources

The objective of this tool is to provide a set of methods to register users, devices and resources in the simplest way possible. These methods, that conform an API, will be used by the final user to get registered through a specially designed user interface. Once the user is registered, the Registration tool will interact later with the corresponding platform resource directory to properly register devices and with the user's directory, assisted by the SocIoTal's Security Framework, to check the user credentials. Figure 2 presents the initial scheme for the Registration Tool.

2) Discovering people, devices and resources

To improve the sharing information process between users, different types of discovery will be developed. Through the SocIoTal Discovery Tool users will be able to discover other users by using different filters such as geolocation, community to which they belong, etc. In addition, the tool will allow the users to discover devices/resources filtering by different properties such as geolocation, entities, attributes, etc. An initial scheme for the Discovering Tool is depicted in Figure 3.

3) Community Creation

Based on the conducted comprehensive set of interviews, surveys and workshops, as one of the most important barriers in the IoT is recognized user's acceptance of the fear of loosing data control and personal privacy. Users want to share information without information leakage and having the control of their data in every moment.

In order to achieve that, the Community Creation Tool allows creating groups where the information is only shared among authorized members thanks to the SocIoTal Security framework. Functionalities such as owner assignation, add/remove user/resources/storage, modify security policies, etc. are provided by the tool. An initial scheme for the Community Creation Tool is presented in Figure 4.

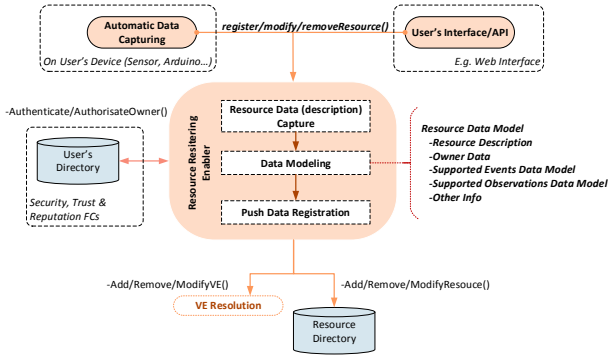


Figure 2. SocIoTal Registering Tool initial diagram

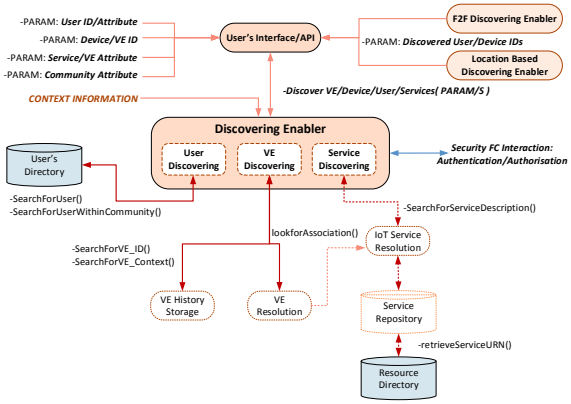


Figure 3. SocIoTal Discovery Tool initial diagram

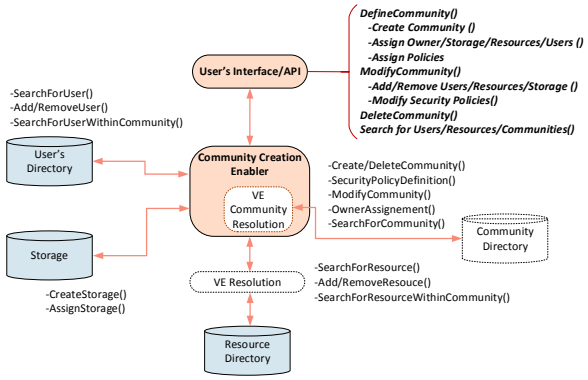


Figure 4. SocIoTal Community Creation Tool initial diagram

4) Evaluating Mood of the city

"Mood of the city" is a concept defined in SocIoTal project that offers citizens a method to assess their mood and share it with other citizens. There are previous methods for evaluating peoples' mood [5] and happiness [6], [7]. This use case tries to provide a joint metric that will help citizens to measure mood in the city by introducing contextually different parameters [8] to previous work.

This trial will allow the evaluation of the mood of the city enabler that offers to the users a method to assess their mood based on data entered (i.e. picture of their face and answers to the specific question) as well as based on current environmental data collected in the city using sensor devices

a) Collecting environmental data

Environmental data, i.e. humidity and temperature are summoned from Ekobus device [9], a shield sensor board attached to a rooftop of a public transportation vehicle in Novi Sad. These data are collecting and processing in the local database, on every few minutes. Also, users can access these data through an Android mobile application.

b) Collecting user's mood data

This main functionality of this use case is mood detection from the users' facial expression as well as collection of the users' happiness index by using happiness index questionnaire. Using mobile application's camera, user detects his mood and then populates questionnaire, with a set of questions commonly used in scientific community for evaluating peoples' happiness.

c) Computing mood of the city index

This use case is focusing on several functionalities offered to the user in order to compute the final mood which is going to be presented to the user. Data gathered from all users are then combined and used to compute mood of the city index. Inputs are scaled to a predefined maximum impact factor that each parameter has and then final value is given to the user as overall summation ratio of all inputs.

5) Evaluating Elevator Supervisor

One of the challenges tenants have from time to time in building maintenance, is elevator repairs. There is no efficient method to know which distance elevator reaches between scheduled inspections, thus this number can be significantly different in the same time intervals. To provide better insight into these numbers, as well as to provide automatic detection of elevator malfunction elevator supervisor use case is proposed.

This trial will evaluate elevator supervisor deployment that enables tenants to monitor history of repairs, elevator distance travelled between inspections and to signal when a new repair is required, as well as to detect malfunction. SocIoTal Web application portal can be used to add new users that can monitor elevator condition, schedule next inspection depending on number of travelled kilometers, put alarms for elevator jams and inspections.

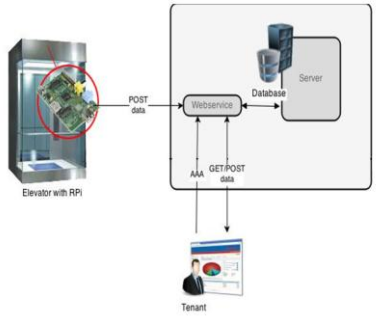


Figure 5. Elevator Supervisor field trial

In this scenario two use cases are identified:

a) Detecting elevator's travelled distance

This use case is based on a set of HW and SW tools that enable monitoring of elevator travelled distance. Raspberry Pi (RPi) device [10] with accelerometer is used to detect movements and after postposing of the signal using low pass filter a number of travelled meters is calculated and sent to the SocIoTal Web application. Users can track travelled distance of the elevator in order to schedule an elevator inspection.

b) Detecting elevator malfunction

This use case is used to automatically detect elevator malfunction by using accelerometers' and PIR sensor [11] data attached on the RPi board. If detected, details about malfunction are sent to the SocIoTal Web portal. User can create notifications in case of malfunction in order to be alerted in case of emergency.

B. Field Trial Evaluation Methodology

The main objective of this process is to gather, from every trial's set of final users, the results of testing each tool/s and/or enabler/s involved, using the mechanisms defined through this text. The target groups are attracted from local events and workshops, and through these activities the project interacts with them and gathers feedback regarding the potential usage of the project outputs as well as new requirements, potential additional functionalities, features that the SocIoTal solution should provide and also a rich evaluation in the different evaluation phases. The target groups considered at the moment are: end users as citizens not directly involved in technology; citizens with a higher level than "user-level" knowledge but without being experts; and service developers as a group involved in the creation of high value services for citizens. This last group is expected to provide more technical feedback which will help the project to capture new requirements, new possible features, technical bugs or malfunctions.

The enablers and tools to be developed during the project life will be evaluated in different phases and within each step a target group (or several) will be approached. Firstly, the first version of the enabler will be evaluated internally within the project partners with the purpose of fixing first bugs and malfunctions. End users will evaluate a following version of the enabler, those will be citizens and developers selected from workshops, and technology savy people interested in the project. In order to obtain a complete evaluation that aims to have a final and stable version of the tool, questionnaires will be distributed to the different target groups during and at the end of the experiment. Also, an email communication channel will be provided in order to report bugs, malfunctions, suggestions, etc.

During the evaluation a set of different KPIs (Key Performance Indicators) will be tested. These indicators involve different aspects such as number of evaluators, % of failures during the execution of the tools, process performance time, look and feel, usability, correct security processes executions, accuracy of data, user trust, energy and data spent, etc.

III. PILOTS

The selected SocIoTal service pilots fulfill two of the most important achievements of the project: to bring to the final users the platform, tools and the developed enablers together, including the mechanisms designed to engage and enrol them and, as a result, to collect the feedback related to their experience, together with performance and acceptance of the SocIoTal innovations. This way, a proper performance of the selected pilots will conform the best method to evaluate SocIoTal as a whole. In next paragraphs, the pilots to be deployed in Novi Sad and Santander will be described and then a summary of the evaluation process will be presented.

A. Novi Sad Pilots

Novi Sad pilot will implement two different pilot trials: Elevator Supervisor and Mood of the city. Elevator supervisor field trial is depicted in co-creation workshop with citizens held in Novi Sad, as one of the use cases that participants were showing the most interest for. Mood of the city is a novel concept calculated using a set of scalable inputs collected from the citizens, i.e. citizens' mood and environmental data collected from sensors.

1) Elevator Supervisor

This field trial is deployed in a resident building where sensor and application provide elevator malfunction detection, data access control and notification for multiple users as well as history track and information about previous repairs and inspections. The main goal of the application is to enable history track of previous repairs and to signalize when the elevator maintenance should be done after certain travelled distance. The pilot is deployed using Raspberry Pi [10] board with GPRS interface, 3-axis accelerometer to detect elevator movements and PIR sensor for presence detection. Movements' value are logged at the SD card and evaluated in a real-time in order to detect certain behaviour. Data is saved to a web application which limits access to authorized users only. Data access control and notifications for multiple users (e.g. tenants, company responsible for repair, etc.) enables history track, information about previous repairs and scheduling of future repairs.

2) Mood of the city

This pilot will enable computation of a Mood of the city defined herein as a scaled metric, derived from contextually different entities that influence people happiness and mood. Final Mood of the city value is computed using aggregated users' data, i.e. users' mood detected from an image, answers to a subjective happiness questionnaire [13], users' selections from a predefined mood list and environmental data.

It is worth noting that responses to the happiness measure cannot be attributed to respondents' current mood [13]. In addition, as final value will depend of a number of users posted their data, we have included well-known parameters that are proven to influence peoples' mood: environmental parameters, i.e. temperature and humidity [8]. Before detecting emotions using a camera, the face region is extracted from the image. From the face, mouth area is cropped and then classified using Fisherface algorithm [14] trained on a Yale dataset [15] increased with a series of custom labelled images. The algorithm is capable of detecting three types of emotions: happy, sad and normal. Implementation is done using OpenCV android library [16]. Temperature and humidity are summoned from EkoBus [9] devices; sensors attached to a moving public transportation vehicles. All these are used as an input to build the final mood of the city index.

B. Santander Pilots

In Santander Sharing Information and the Enabling Santander are two pilots extracted from ideas shared by citizens through the Santander City Brain platform [17]. In the case of the Enabling Santander, this idea was led by an external group of developers and, through the first Santander IoT Meetup [18] it was incorporated to SocIoTal as one of its scenarios with the objective of building the application over the tools provided by SocIoTal.

1) Sharing Information

Sharing Information pilot is shaped as a service that will provide citizens with a platform that will allow them sharing their own data (from their devices) only with people to whom they give permission, within a secured and trusted environment.

This pilot will use mainly the Registering users/devices, Community Creation and Discovery tools [19] but can be easily modified to include other tools developed within SocIoTal.

Users will firstly register themselves against the platform through the SocIoTal Registering Tool [19].

Once the user is registered they will be able to register their devices. All devices capable to send their measurements to a server are susceptible to be registered within the SocIoTal platform. Users can share information from different sources, for example, information gathered by the sensors included within their smartphones or tablets. In the case that they are more interested in technological DIY gadgets, they could build their own devices such as weather stations through Arduino or Raspberry Pi boards and different sensors and in an easy way program them to send their observations to the recipient platform. Once the Community has been created, members will be able to access information produced by the rest of members in the Community through the SocIoTal User Environment. To request information about users, devices or observations within the community the users will fill the needed information to describe what are they looking for, and the SocIoTal Discovery tool will be in charge of discover all data related to the request that users are allowed to access. In addition to this, citizens will be able to receive information following a pub/sub pattern, subscribing their profile to the different resources.

2) Enabling Santander

Enabling Santander pilot will provide disabled citizens with an application to go from one place to another in the city, avoiding barriers along their journey (works, road closed, narrow sidewalk, etc.). This application, proposed by a group of external developers in the city, will be built over some necessary SocIoTal tools.

The application will make use of the SocIoTal Registration enabler in order to register the users and give them the authorization to access the platform and use the services

Within the application, the route calculation service will make use of the SocIoTal Discovery Tool. When the user requests a route, the route calculation algorithm will need data about the barriers that can cause a problem for a disabled person. These events will be requested through the Discovery Enabler. Also, a user through the application could request data about the accessibility of a concrete area by only selecting the area they want to analyze.

C. *Evaluation Methodology*

It can be highlighted that the evaluation process has started from the beginning of the project when the use cases were selected since they were extracted from ideas gathered within workshops with citizens or from platforms where users can upload ideas and vote them. From that use cases a smaller group of ideas were selected and adapted to turn into pilots.

Within the pilot's evaluation process two main methodologies and tools are considered: questionnaires and qualitative interviews from target groups collected during workshops where the services are presented to the end users, and real life testing. The questionnaires are composed with the aim to collect useful information on improvements of the presented pilot as well as end users (i.e. citizens) satisfaction with the current implementation. The data is going to be collected during workshops where service is presented to the end users and their feedback collected for the further pilot improvement. In addition, the description of a set of different test cases are be provided to check the correct operation of the functionalities of the pilot. These test cases will be focused in different aspects such as API usability and performance, correct operation of the different tools forming the bases of the functionalities, security, look and feel, data accuracy, etc. Finally, all feedback collected will be used for driving corrective and improvements measures for the platform. Evaluation process is shown in Figure 6.

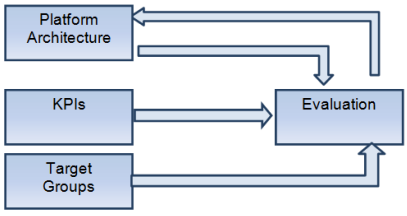


Figure 6. Evaluation process

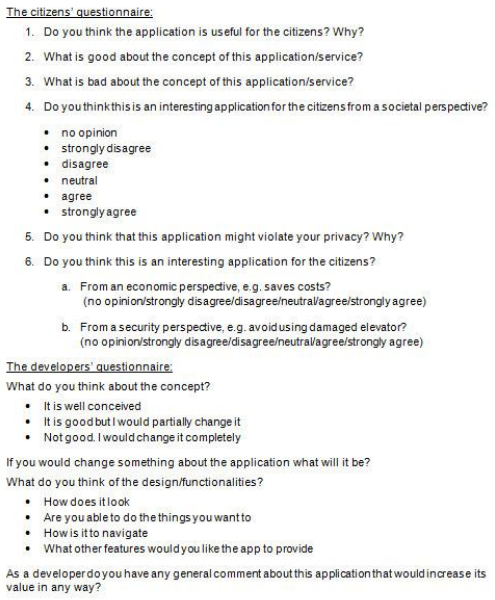
IV. Key Performance Indicators and evaluation questionnaires

In this section KPIs and questionnaires for evaluating the field trials are presented. The data is going to be collected during workshops where service is presented to the end users and their feedback collected for the further pilot improvement.

TABLE I. Field trials' Key performance indicators

KPI Id:	KPI title:	Definition
SocIoTal Context Management Tools Key Performance Indicators (KPI)		
001	**Number of evaluators**	Number of people from the target groups that will evaluate the corresponding tool/API.
002	**Tools Usability**	Usability measures the grade of simplicity, adaptability and functionality perceived by users when they perform the corresponding tests through the provided tools
003	**General Tool crashes ratio**	Percentage of tool crashes during its usage, due to APIs malfunction coming from issues out of SocIoTal development (selected platform crashes, communication links failures, etc.).
004	**Process performance time**	Time the user takes to execute (prepare and send the request and receive the response) the corresponding procedure
005	**API Usability**	API usability measures the grade of simplicity, adaptability and functionality perceived by users (mainly developer – geek users) when they perform the corresponding tests through the provided APIs
006	**Failed process execution ratio**	Percentage of errors occurred during the execution of the analysed process
007	**Procedure (API) response time**	Related to the time the procedure API takes to retrieve Ok once it's been called
Community Creation tool Key Performance Indicators (KPI)		
008	**Usability**	Usability measures the grade of simplicity, adaptability and functionality perceived by users when they perform the corresponding tests
009	**CM Tool crashes ratio**	Percentage of tool crashes during its usage
010	**Process performance time (User Interface)**	Time the user takes to execute (prepare and send the request and receive the response) the corresponding procedure
011	**Process performance time (API response time)**	Time the user takes to execute (prepare and send the request and receive the corresponding procedure response)
Mood of the city Key Performance Indicators (KPI)		
001	**Number of evaluators**	Number of people from the target groups that will evaluate the Mood of the city enabler.
012	**Usability**	Usability measures the grade of simplicity perceived by the users when they use Mood of the city application.
013	**% Application crashes**	Percentage of application crashes during the usage of the Mood of the city enabler
014	**Process performance time**	Time the user takes to provide a data (current image of themselves and answers to the specific question)
015	**% Facial expression detection accuracy**	Percentage of success Facial expression detection accuracy
016	**% Environmental data accuracy**	Percentage of accuracy of environmental data
017	**Look and feel**	This KPI tries to measures the look and feel perceived by the users when they receive the results.
Elevator supervisor Key Performance Indicators (KPI)		
001	**Number of evaluators**	Number of people from the target groups that will evaluate the elevator supervisor enabler.
017	**Look and feel**	This KPI tries to measures the look and feel perceived by the users when they receive the results.
018	**Usability**	Usability measures the grade of

		simplicity perceived by users when they use elevator supervisor application.
019	**% Application crashes**	Percentage of application crashes during the usage of the elevator supervisor enabler
020	**% Malfunction detection accuracy**	Percentage of success of malfunction detection accuracy
021	**% Travelled distance calculation accuracy**	Percentage of success of travelled distance calculation accuracy



The citizens' questionnaire:

1. Do you think the application is useful for the citizens? Why?
2. What is good about the concept of this application/service?
3. What is bad about the concept of this application/service?
4. Do you think this is an interesting application for the citizens from a societal perspective?
   - no opinion
   - strongly disagree
   - disagree
   - neutral
   - agree
   - strongly agree
5. Do you think that this application might violate your privacy? Why?
6. Do you think this is an interesting application for the citizens?
   a. From an economic perspective, e.g. saves costs? (no opinion/strongly disagree/disagree/neutral/agree/strongly agree)
   b. From a security perspective, e.g. avoid using damaged elevator? (no opinion/strongly disagree/disagree/neutral/agree/strongly agree)

The developers' questionnaire:

What do you think about the concept?

- It is well conceived
- It is good but I would partially change it
- Not good. I would change it completely

If you would change something about the application what will it be?

What do you think of the design/functionalities?

- How does it look
- Are you able to do the things you want to
- How is it to navigate
- What other features would you like the app to provide

As a developer do you have any general comment about this application that would increase its value in any way?

Figure 7. The evaluation questionnaire

Some of the KPIs are common (the same) for a few scenarios, but they are provided for all pilots for readability. Some KPIs have the same name (like KPIs 002 and 008: Tools Usability) and perform the same measures, but evaluate different tests (for example KPI 002 performs test: Create a new SocIoTal user/identity, while KPI 008 performs test: Using the community Management Tool to create/update/modify/delete community.

V. CONCLUSION

In this paper we have presented the scenarios selected for the field trials and pilot deployment, together with the evaluation methodology, including relevant KPIs. The trials are deployed within the project in order to test the different tools and enablers developed within SocIoTal. The purpose of field trials and pilots is to test the developments over real environments, with real users, facing all the constraints and limitations that a complex society can pose in these kinds of trials. Appropriate test cases and evaluation methodologies are described. Within the pilot's evaluation process different methodologies and tools are considered: questionnaires and qualitative interviews from target groups collected during workshops where the services are presented to the end users, and real life testing. A summary of the methodology followed in the project to evaluate the different pilots described above in terms of acceptance by the final users and correct operation of their functionalities is given.

ACKNOWLEDGMENT

This paper describes work undertaken in the context of the SocIoTal project (http://sociotal.eu/). The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement n° CNECT-ICT- 609112.

REFERENCES

[1] M. Victoria Moreno, José L. Hernández, Antonio F. Skarmeta, Michele Nati, Nick Palaghias, Alexander Gluhak, Rob van Kranenburg, "A Framework for Citizen Participation in the Internet of Things", Pervasive Internet of Things and Smart Cities (PitSac) whorkshop, 2014.

[2] Bernabe, J. B., Hernández, J. L., Moreno, M. V., & Gomez, A. F. S. (2014). Privacy-Preserving Security Framework for a Social-Aware Internet of Things. In Ubiquitous Computing and Ambient Intelligence. Personalisation and User Adapted Services (pp. 408-415). Springer International Publishing.

[3] Nenad Gligoric, Srdjan Krco, Ignacio Elicegui, Carmen López, Luis Sánchez, Michele Nati, Rob van Kranenburg, M. Victoria Moreno, Davide Carboni , "SocIoTal: Creating a Citizen-Centric Internet of Things", ICIST 2014, 4th International Conference on Information Society and Technology

[4] EU FP7 Internet of Things Architecture project, http://www.iot-a.eu/public, last accessed 28/11/2014

[5] Mit Mood Meter, [Online]. Available: http://moodmeter.media.mit.edu/, accessed 31.10.2014

[6] Anna Maffioletti,Agata Maida, Francesco Scacciati, "More Terminological and Methodological Problems in Measuring Happiness, Life Satisfaction and Well-Being: Some First Empirical Results", The Pursuit of Happiness and the Traditions of Wisdom SpringerBriefs in Well-Being and Quality of Life Research 2014, pp 13-21

[7] Peter Hills, Michael Argyle, The Oxford Happiness Questionnaire: a compact scale for the measurement of psychological well-being, Personality and Individual Differences Volume 33, Issue 7, November 2002, Pages 1073–1082

[8] Howarth, E., & Hoffman, M. S. (1984). A Multidimensional Approach to the Relationship between Mood Weather. British Journal of Psychology, 75 and (1), 15-23

[9] CITI-SENSE FP7 EU project. Official Deliverable D8.2., "Pilot studies platforms"

[10] Raspberry-Pi platform [Online]. Available: http://www.raspberrypi.org/ , Feb. 7,2014

[11] PIR motion sensor [Online]. Available https://learn.adafruit.com/pir-passive-infrared-proximity-motion-sensor

[12] SOCIOTAL FP7 STREP EU project. Official Deliverable D6.1, " D6.1 Report on first year community interactions and detailed dissemination strategy"

[13] Sonja Lyubomirsky, Heidi S. Lepper A Measure of Subjective Happiness: Preliminary Reliability and Construct Validation, Social Indicators Research, February 1999, Volume 46, Issue 2, pp 137-155

[14] PN Belhumeur, JP Hespanha, D Kriegman, Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, Pattern Analysis and Machine Intelligence, IEEE Transactions on 19 (7), 711-720

[15] YALE Dataset Athinodoros Georghiades, Peter Belhumeur, and David Kriegman's paper, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose", PAMI, 2001

[16] OpenCV [Online], Available: http://docs.opencv.org/index.html

[17] Santander City Brain. [Online]. Available: http://www.santandercitybrain.com/ , [Nov. 13, 2014]

[18] IoT Santander Meetup. [Online]. Available: http://www.meetup.com/IoT-Santander/

[19] SOCIOTAL FP7 STREP EU project. Official Deliverable D5.1, " D5.1 Trial and pilot specifications" Report on first year community interactions and detailed dissemination strategy"

PyTabs: A DSL for simplified music notation

Miloš Simić, Željko Bal, Renata Vaderna, Igor Dejanović
Faculty of Technical Sciences, Novi Sad, Serbia
{milossimicsimo, zeljko.bal}@gmail.com, {vrenata, igord}@uns.ac.rs

***Abstract*—In this paper we present pyTabs – a DomainSpecific Language (DSL) for simplified music notation. In pyTabs it is possible to describe a composition that consists of multiple sequences which can be specified in the form of a tablature or chord notation. One notable feature of pyTabs is the capability to play a musical piece written in it. We describe some major issues in simplified music notations (tablatures and chords) and propose a solution implemented in pyTabs project as a way of standardizing them into a formal language. puTabs is a free and open source project implemented in python programming language.**

It is available at: https://github.com/E2Music/pyTabs

I. INTRODUCTION

pyTabs is a DSL for simplified music notation and composition description. Domain-Specific Languages (DSLs) [1], in contrast to general-purpose languages (GPL), offer, through specific notations and abstractions, the power of expression focused on, and usually restricted to, a particular problem domain. DSLs are classified by Martin Fowler [2] on the basis of their construction as:

- External DSLs - built from scratch, with their syntax carefully tailored for the domain in question. Often called little languages.

- Internal DSLs - built on top of an existing GPL, extending their syntax to add support for domain-specific constructs. pyTabs is an external DSL. Another classification of DSLs is [3]:

- Technical DSL - used by programmers and

- Non-technical or application domain DSL used by non-programmers. pyTabs is an application domain DSL (sometimes also called business DSL or vertical DSL) meant to be used by music players/compositors.

This language is developed for the people who are not experts in writing and/or playing music. Because tablature and chord notations are relatively simple and intuitive, they are easy to learn and understand.

Therefore a lot of people who decide to start playing music usually first start with them. pyTabs language extends the basic form of chords and tablatures, trying to enrich them and fix some of the major problems in these notations. Also, at the same time it tries to stay easy and intuitive to the people who are used to the standard notations. pyTabs goes a little bit further, and allows playback of compositions written in this way. Fixing major problems with the standard notation, composition playback and also knowing that more than 800 000 songs are available in tablature notation online [4], means that pyTabs can really help with learning.

The paper is structured as follows: Section 2 describes the tablature notation; In Section 3 we give the current problems with the existing tablature notation; Section 4 gives a description of the pyTabs language, while section 5 describes the architecture of the project; Section 6 describes the tools that were used in the project; In Section 7 related work has been presented. In Section 8 we conclude the paper.

II. ABOUT TABLATURE NOTATION

Tablature [5] (or tablature, or tab for short) is a form of musical notation indicating instrument fingering rather than musical pitches (figure 1). While standard notation represents the rhythm and duration of each note and its pitch relative to the scale based on a twelve tone division of the octave, tablature is instead operationally based, indicating where and when a finger should be placed to generate a note, so pitch is denoted implicitly rather than explicitly.

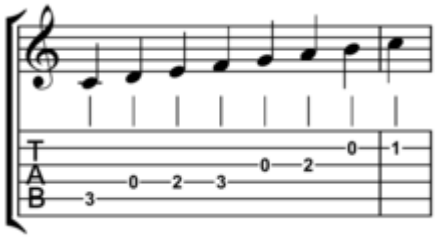


Figure 1: standard musical notation above, tablature notation below

III. CURRENT PROBLEMS IN TABLATURE NOTATIONS

There are two major problems with the current tablature notation. The first one would be a visual problem. Tablatures have not been fully standardized so far and anyone can write them as they like, making their own variations of the notation. This brings some visual problems, especially when a line is not well-formed. Do we want to play more than one note, or just one note per string? What if it is a two digit number on one line and a one digit number on another, how many dashes should there be between the numbers, and so on (figure 2).

```
e|---0---1---3---
B|---0---1---0---
G|-----10-2---0---
D|---2---3---0---
A|---2---3---2---
E|---0---12---3---
```

Figure 2. Tablature visual problem

The second problem is that there is no standard way of specifying the note duration in a tablature or chord notation. It is usually implicitly inferred by the player

who knows the rhythm of the song, but it is impossible for someone to play the song properly without first knowing the rhythm.

IV. PYTABS LANGUANGE

pyTabs language is designed to improve quality and to bring some form of standardization to the tablature notation.

A. Tablature

A tablature in a textual format consists of one or more rows (6 for a standard guitar, 4 for a standard bass guitar and so on). Each row starts with a symbol representing a row (i.e. string letter for a guitar tab) and contains a number of symbols, usually representing notes, divided by one or more dashes ("-"). A break is represented by a single dash (figure 3).

```
e|---0---1---3---
```

Figure 3. Tablature row example

All the symbols are organized into columns which represent the notes that should be played in an instance of time. If one symbol in a column consists of more characters than the others, the other columns must be padded with dashes so that the symbols that follow can be placed in the same column. In order to create a formal language that can be parsed with a text parser we've accepted a set of rules that are common to most tablature formats. Figure 4 shows an example of a tablature written in pyTabs.

```
e|-0----10-3-||
B|-0----1--1-||
G|-12pm----6-||
D|-2----9--0-||
A|-2----3--2-||
E|----------||
```

Figure 4. Tablature example in pyTabs

The main problem with parsing a tablature in this format is the fact that the interpretation of a symbol is dependent on the length of the other symbols in the same column. A dash could be interpreted as a break or as a padding. Thus a linear text parser cannot be used in this case. The solution to this problem implemented in pyTabs is to parse the tablature column by column taking into account the length of every symbol in a column. This is achieved by recognizing the tablature as a set of rows and later recognizing the individual symbols column by column while removing the padding dashes. The number of padding dashes in each row is determined by the longest symbol in the current column.

Since there are different tablature notations for various instruments and they all share some common rules it was useful to extract the logic about parsing a tablature into a generic tablature parser. Parsing rules that are specific to each instrument could later be defined in a concrete implementation.

After parsing a tablature in this way a set of row models is obtained. Each row has a symbol denoting the row and a set of symbols that represent the contents. The semantics of the individual symbol can then be determined by the row mark and the note symbol itself. The generic tablature processor performs the parsing column by column and delegates the individual symbol recognition to a note processor provided at initialization. A note processor takes two arguments (a row mark and an actual note symbol) and returns an object representing the note semantics (i.e. for a guitar a row mark 'e' and a symbol 0 are translated into 'e' string with fret 0). The objects returned by a note processor are then packed into container objects column by column (where each container object represents a time instance) which are then packed into a resulting list (the container object type can also be provided on initialization). At the end the list of container objects is returned as a result.

This way the only thing required for creating a tablature processor for a new instrument is to create a note processor for that instrument and pass it to the generic tablature processor. The generic tablature parser uses textX (see section 6.1) to obtain a tablature model based on a generic tablature grammar and passes it to the generic tablature processor that returns a resulting list. The processor can be used independently which is the case when the tablature model is obtained through a textX composition model. The guitar note processor for example uses a separate textX grammar for individual note parsing. A significant shortcoming of tablature notation is the lack of a standard way to specify the note duration. This makes it impossible to properly render a song written in it, so since one of the main purposes of *pyTabs* is to play the music according to tablatures, a standard way of specifying the note duration had to be defined. The solution implemented in *pyTabs* is to add a row marked with the letter 'R' (for rhythm) which contains numbers denoting the column duration (i.e. 4 for 1/4, 8 for 1/8 and so on). This additional column behaves the same way as the others (separated and padded with dashes) and can be recognized by a generic tablature processor and therefore is easily implemented in a note processor.

B. Chords

Chord in music, by definition [6] is *three or more musical notes played at the same time*. This group of tones usually has a name given by the major tone in the sequence and by that name musicians know which chord exactly to play. This is shorter to write and easier to remember. This notation usually represents a rhythm part of a composition. Figure 5 shows various chords.

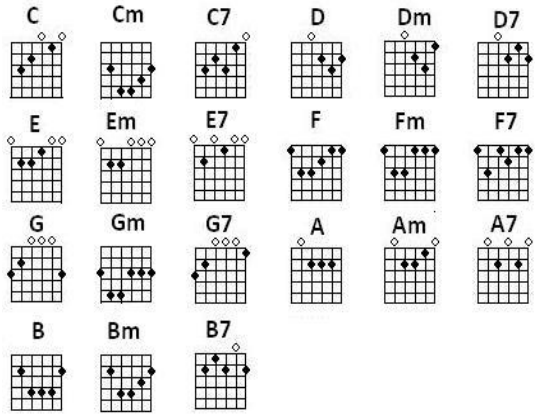


Figure 5. Various chords diagrams

In *pyTabs* every chord construction starts with one of twelve basic tones. These tones are C, C#, D, D#, E, F, F#, G, G#, A, A#(B), H(B), where tones in parentheses represent the difference between North American, and European naming conventions. After the main chord tone may come a *number* that represents the specific tone from a scale that has been added to the chord (C5, C7, …) or a *decoration* that changes the scale of the chord (m, maj, sus, …). After the decoration there may also be a number that is added from the scale (Cmaj7, Csus4). These two parts can be combined into one and/or separated with a "/" sign (A/G chord) to build more complex chords. If there is nothing after the main tone name, then that is a major chord. To fix the lack of a way to specify the note duration in the standard notation, every chord comes with its time duration (whole tone, half tone …). For this purpose chords grammar is extended with time duration inside parentheses.

If a chord is "G minor seven" and we want that chord to continue for the whole tone, that construct is Gm7(4).

Rhythm sections can be on a pause for a while and then start playing again, or they can be constructed into a *riff* which is a repeating pattern of chords and pauses. For this reasons the chords grammar is also extended with pause parts. Pause parts are represented by brackets with a number that represents the pause duration ([8],[4],[2]…).

C. Composition

Composition is divided into five parts and its role is to create a composition model ready to be played. The first part is some basic data about the song: author, name, tempo and beat. The second part is the import section where the name and location of the *sound font* that contains the sound samples are given in form of key-value pairs. This is usually a list of key-value pairs because more than one instrument can play in a composition. Third part is the sequence list. Every sequence starts with a '*sequence'* keyword followed by the type (guitar-rhythm, guitar-solo, bass, drums, etc.), the name and the contents of the sequence. The contents hold tablature or chord elements. After the sequence list comes the segment list. Its job is to connect the sequence name to the instrument name in order to know which sound font is played by which sequence. Segment part begins with a '*segment'* keyword followed by the segment name and a list of *sequence* name and *import* name pairs separated with a ":". This part represents parts of the song (Chorus, Solo, Bridge, Verse …). Last part of a composition is a *timeline.* Its job is to connect segments into one song. It starts with a keyword '*timeline'* and inside curly brackets we put a list of segment names in order that we want them to be played in, separated by a "," (Intro, Verse, Chorus …). Figure 6 shows an example of a song written in pyTabs language.

```
[
    Name "Dim na vodi"
    Author "Tim1"
    Beat 4/4
    Tempo 120
]

import
bass "instruments/Soundfont BassFing.sf2"
guitar "instruments/Saber_5ths_and_3rds.sf2"

sequence guitar-solo bass_tabs
{

R|-8-8-8-8---8-8-8-8-8-8-----8-8-8-8---8-8-8-2-|
G|----------------------------------------------|
D|----------------------------------------------|
A|----------------------------------------------|
E|-0-0-0-0-0-0-0-0-0-3-2-1-0-0-3-3-5-5-3-3-0-|

}

sequence guitar-rhythm guitar_chords
{
    A(4) B(4) C(4) D(4) E(4) F(4) G(4)
}

segment Chorus
{
    bass_tabs : bass
    guitar_chords : guitar
}

timeline
{
    Chorus
}
```

song meta data

import section

sequences section

song segments section

song timeline

Figure 6. An example of a composition written in pyTabs

V. ARCHITECTURE

pyTabs is composed of two parts: 1) the editor which is responsible for editing, syntax highlighting and sending a model to the engine and 2) the engine which knows how to process the model data.

A. Editor

The editor is developed using QT library and PySide wrapper for Python (details in the next chapter). The main component is SyntaxHighlighter.

Its job is to highlight the language syntax, and also to help user with writing. This is accomplished by a regular expression and/or with a list of reserved words connected with color.

This task is done by HighlightingRule class which maps the reserved words and color. Since, this class is connection with reserved word and color, it must be created as many different instances as there are different parts of the language.

Figure 7 presents the editor user interface, with a composition example and text highlighting.

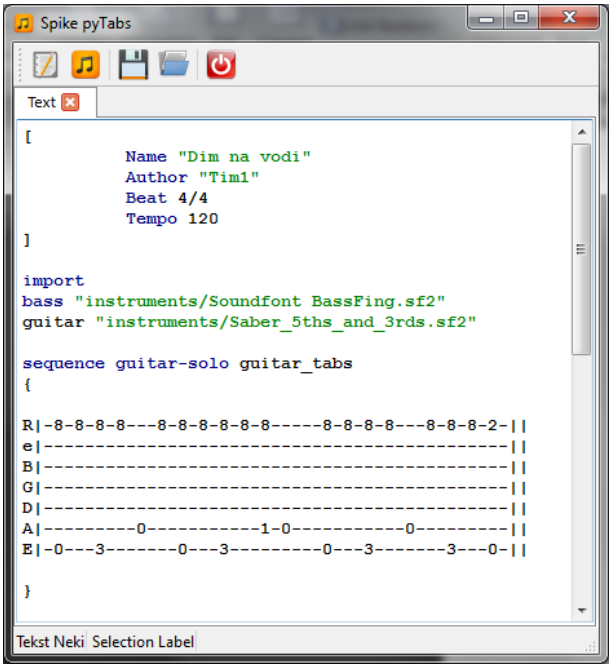


Figure 7. Editor window

B. *Engine*

pyTabs engine is separated in two major parts, *Composition* and *Player*. Figure 8 presents *pyTabs* engine class diagram.

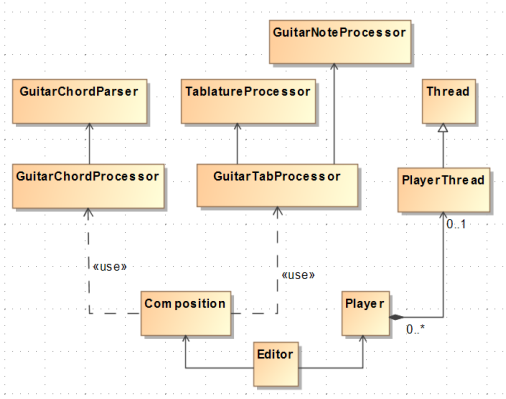


Figure 8. Engine class diagram

The player module uses FluidSynth (via a wrapper provided by the Mingus library, see sections 6.2 and 6.3) to play the composition based on the composition model. The tracks are played sequentially. Each segment in a track is played in a separate thread simultaneously since FluidSynths play methods are blocking.

The composition module uses textX and composition grammar to parse a composition description and obtain a composition model. It registers textX object processors for sound font imports and song sequence processing. The first one organizes the instrument imports into a map (instrument name to sound font file path) and the second one processes the sequences based on their input type (tablature, chords etc.). The value of each sequence is replaced by a Mingus Track object that is created using a specific input type processor. The Track object is created based on the information in the sequence model (note pitch, duration etc.). The resulting model is suitable for later use with the Mingus library (i.e. playing the song using FluidSynth).

VI. TOOLS

A. *TextX*

textX [7] is a meta-language for building Domain-Specific Languages (DSLs). From a single language description (grammar) textX will build a parser and a meta-model (a.k.a. abstract syntax) for the language. textX is used for language grammar construction, and creating model from it.

B. *Mingus*

Mingus [8] is a package for Python used by programmers, musicians, composers and researchers to make and investigate music. Some important features:

- The Note class: can keep track of octaves, dynamics and effect and also allows you to compare Notes: eg. Note("A") <= Note("B")and convert to and from Hertz.
- Data structures that group notes together in blocks of notes (NoteContainers), Bars, Tracks, Compositions and Suites.
- A MIDI sequencer which uses the container objects and can send timed MIDI messages to an output function. Support for fluidsynth (a software MIDI synthesizer), so that objects can be played in real-time.

In *pyTabs* Mingus is used for representing and grouping the notes using the classes in mingus.containers package (Note, NoteContainer, Track, etc.).

C. *Fluidsynth*

FluidSynth [9] is a real-time software synthesizer based on the SoundFont 2 specifications and has reached widespread distribution.

SoundFont is a brand name that collectively refers to a file format and associated technology designed to bridge the gap between recorded and synthesized audio, especially for the purposes of computer music composition. SoundFont [10] technology is an implementation of sample-based synthesis.

Sample-playback-based MIDI synthesizers use wavetables to define the base samples that are used to render their MIDI files. MIDI files in themselves don't contain any sounds, rather they contain only instructions to render them, and consequently rely on the wavetables to render such sounds correctly. SoundFont-compatible synthesizers allow users to use SoundFont banks to augment these wavetables with custom samples to render their music. The fluidsynth wrapper is used to play the composition using sound fonts based on the composition model.

D. QT and PySide

Qt [11] is a cross-platform application framework from Qt Software (owned by Nokia). It features a large number of libraries providing services like network abstraction and XML handling, along with a very rich GUI package, allowing C++ developers to write their applications once and run them unmodified in different systems.

PySide [12] aims to provide Python developers access to the Qt libraries in the most natural way. In pyTabs, PySide is used to create user interface.

VII. Related work

Tablature notation is an alternative to standard musical notation. It is popular among people who start learning to play a musical instrument, especially guitar.

Guitar pro [13] is the most popular commercial software, but it is not suitable for beginners.

Tabledit [14] is a little bit simpler but also a commercial tool and not so beginner friendly.

Still these tools are not meant for the people who are learning how to play an instrument.

VIII. Conclusion

In this paper we have presented the *pyTabs* DSL for tablature notation. We have also presented one possible solution to fixing two major problems in current notations, by adding duration to tablatures and chords and formatting the tablature lines in such a way that we keep the simplicity and intuitiveness currently available in the notations to which people are accustomed.

We have presented a possibility of connecting different notations in a composition, with the ability to playback the compositions written in this way using mingus, fluidsynth and soundfont standard.

In the further work we plan to add more instruments andresearch the way of their integration. Also we plan to addthe ability to generate the standard musical notations from pyTabs and vice versa.

References

[1] Van Deursen, A., Visser, J.: Domain-specic languages: an annotated bibliography, ACM SIGPLAN Notices , vol. 35, pp. 26-36, 2000

[2] Fowler, M. Domain-Specific Languages Addison-Wesley Professional, 2010

[3] Völter, M. DSL Engineering: Designing, Implementing and Using Domain-Specific Languages, 2013

[4] Ultimate-guitar website, http://www.ultimate-guitar.com/, accessed 5. January 2015.

[5] Tablatures wikipedia article, http://en.wikipedia.org/wiki/Tablature, accessed 9. January 2015.

[6] Elpin Systems, http://www.elpin.com/tutorials/musicalchord.php, accessed 5. January 2015.

[7] textX project page, https://github.com/igordejanovic/textX, accessed 9. January 2015.

[8] Mingus project page, https://code.google.com/p/mingus/, accessed 9. January 2015.

[9] Fluidsynth project page, http://www.fluidsynth.org/, accessed 9. January 2015.

[10] Soundfont wikipedia article, http://en.wikipedia.org/wiki/SoundFont, accessed 9. January 2015.

[11] QT project page, http://qt-project.org, accessed 5. January 2015.

[12] Pyside project page, http://pyside.github.io/docs/pyside/#, accessed 5. January 2015.

[13] Guitar pro page, http://www.guitar-pro.com/en/index.php, accessed 14 January 2015.

[14] Tabledit, http://www.tabledit.com/download/index.shtml, accessed 14 January 2015.

Opportunities of the Internet of Things for Healthcare through Architectural Layers- Architecture and Technologies

Daliborka Mačinković
Health Insurance Fund of Republic of Srpska/Department of Information technology, Banja Luka, BIH, RS
daliborka.macinkovic@teol.net

***Abstract* — The Internet of Things for Healthcare (Healthcare-IoT) includes medical sensor devices as new sources of health information, new technologies and applications for remote diagnostics and monitoring of patients, equipment, drugs. Smart services for healthcare, anywhere and anytime, will bring the future of healthcare services, as a new level of healthcare. This paper considers the Healthcare-IoT through architectural layers and observes the users of the healthcare system in the context of a complex healthcare system which is integrated and whose traditional healthcare services need to be expanded with new functionalities of Healthcare-IoT such as sensing, tracking and monitoring, identification, authentication, automatic data collection. The technologies that should enable an IoT solution for healthcare are presented through architectural layers. The new business model and scenario should include valuable information in the existing healthcare services and choose the technologies that best match the desired model. Sensors and sensor networks, as well as new technologies, need to enable smart objects of healthcare to feel and change the environment, implement activities, communicate in real-time and share information. The data collected should be reliable, safe, processed in real time and analyzed, and they should provide a wealth of intelligence for planning, management and decision-making in the healthcare system. Many challenges for Healthcare-IoT, such as the interoperability during the integration with the inherited systems and electronic health records (EHR), need to be resolved. A holistic approach to designing Healthcare-IoT solutions and interdisciplinary knowledge should remove the gap between individual technology solutions and the integrated healthcare system expanded with IoT functionalities.**

I. INTRODUCTION

This paper considers the opportunities for the Internet of Things in Healthcare with an overview of the existing technologies through the architectural layers and the greatest benefits that can be achieved with Healthcare-IoT.

The main problem which needs to be resolved in this paper is to present the structure and behavior of an integrated healthcare system expanded with the new values of the Healthcare- IoT. New services should be incorporated into the traditional healthcare system by using the most efficient technologies.

The Internet of Things (IoT) is a novel paradigm that is rapidly gaining ground in the scenario of modern wireless telecommunications. The basic idea of this concept is the pervasive presence around us of a variety of things or objects – such as Radio-Frequency IDentification (RFID) tags, sensors, actuators, mobile phones, etc. – which, through unique addressing schemes, are able to interact with each other and cooperate with their neighbors to reach common goals [1].

The phrase "Internet of Things" was coined at the beginning of the 21st century by the MIT Auto-ID Center with special mention to Kevin Ashton (Ashton 2009) [2] and David L. Brock (Brock 2001) [3].

The Internet of Things is a technological revolution that represents the future of computing and communications, and its development depends on dynamic technical innovation in a number of important fields, from wireless sensors to nanotechnology [4].

A successful implementation of IoT solutions requires a suitable infrastructure.

Wireless technologies are arriving in order to ensure the e-health monitoring for patients everywhere and from any given location. Research and development advances in the e-health community include data gathering and transfer of vital information, integration of human machine interface technology into handheld devices, data interoperability and integration with inherited systems and electronic patient records [5].

In the USA, electronic health monitoring has been given the go-ahead by the Federal Communications Commission (FCC). FCC allows the use of allotted frequencies for sensors to control devices wirelessly in the monitoring of health at hospitals and homes, and has also forecast savings [6].

European Space Agency (ESA) has initiated the Digital Video Broadcasting with Return Channel via Satellite (DVB-RCS) [7], technology enabling almost all potential locations - even the most geographically dispersed and isolated ones - to gain access to broadband services using low-cost Satellite Interactive Terminals (SITs). The technology enhanced with the DVB-S2 knowledge is a mature broadband communications technology with comparable implementation and operational costs to the other broadband terrestrial technologies, effectively satisfying the Quality of Service (QoS) requirements of high demanding applications in electronic healthcare [5].

IoT research is faced with the challenges of non-compliance technology in business requirements, as a gap between technology development and business innovation. Many solutions in healthcare are not included in the value chain which reduced their importance and lack of confidence [8].

The problems of the previous research of Healthcare-IoT are the lack of integration of new devices and traditional services. According to WHO [9], the most common result is a noninteroperable abundance of islands ICT. What is necessary is a holistic design that will effectively integrate the scattered devices and technologies into much more valuable services.

This paper is organized into five Sections. After the Introduction, the Section II presents the greatest benefits of Healthcare-IoT which are grouped into tracking and monitoring, identification, authentication, automatic data collection, sensing, and cross-organization integration. The third Section presents integrated Healthcare system extended with IoT for Healthcare through Architectural Layers: Business Layer, Sensor Layer, Network Layer, Services Layer and Application Layer. The Section IV describes the new technologies and communication solutions that will support the development of Healthcare-IoT. The conclusion stresses the importance and opportunities that can be achieved by expanding the traditional services of the integrated health system with new Healthcare-IoT services presented in this paper through architectural layers.

II. BENEFITS WITH HEALTHCARE - IOT

While considering the functionalities of Healthcare-IoT, we can come across a number of concepts that also represent the future healthcare services such as pervasive healthcare (pHelath), ubiquitous healthcare (uHealth)[10], mobile healthcare (mHealth) [11], electronic healthcare (eHealth), telehealth, telemedicine.[12]

The treatment of patients only in medical institutions can be redirected to the treatment at home with complete control, which means that chronic patients will have more freedom and a better quality of life, and that hospitals will get more capacity for emergencies. According to [13], in the coming decade, the model of delivery of healthcare services will be transformed from the present hospital-centric, through hospital-home-balanced in 2020, to the final homecentric.

The Internet of things in the domain of health system includes an increasing number of sensors in the global network. The platform for the Internet of things must enable the processing of new data in real time. User domains differ in the definition of service monitoring of vital functions, administration of appropriate medication depending on the vital parameters of users, monitoring and notification of critical situations. IoT platforms facilitate the sharing of such information among experts in the field of medicine, thus allowing the definition of new procedures for the treatment and diagnosis [14].

One of the prominent areas of research and strategic roadmap of the European Commission for Information Society is the Internet of Things [15].

There are many benefits provided by the IoT technologies for the healthcare domain, and the resulting applications can be grouped mostly into: tracking of objects and people (staff and patients), identification and authentication of people, automatic data collection, sensing, cross-organization integration.[16]

Tracking is the function aimed at the identification of a person or object in motion. This includes both real-time positions of tracking, such as the case of patient-flow monitoring, whose purpose is to improve workflow in hospitals, and tracking of motion through choke points, such as access to designated areas. When it comes to assets, tracking is most frequently applied to continuous inventory location tracking (for example for maintenance, availability when needed and monitoring of use), and materials tracking to prevent left-ins during surgery, such as specimens and blood products [17].

Identification and authentication includes patient identification to reduce incidents harmful to patients (such as wrong drug/dose/time/procedure), comprehensive and current electronic medical record maintenance (both in the in- and out-patient settings), and infant identification in hospitals to prevent mismatching. In relation to staff, identification and authentication is most frequently used to grant access and to improve employee morale by addressing patient safety issues. In relation to assets, identification and authentication is predominantly used to meet the requirements of security procedures, to avoid thefts or losses of important instruments and products [17].

Data collection-Automatic data collection and transfer is mostly aimed at reducing form processing time, process automation (including data entry and collection errors), automated care and procedure auditing, and medical inventory management. This function also relates to integrating RFID technology with other health information and clinical application technologies within a facility and with potential expansions of such networks across providers and locations [17].

Sensing-Sensor devices enable function centered on patients, and in particular on diagnosing patient conditions, providing real-time information on patient health indicators. Application domains include different telemedicine solutions, monitoring patient compliance with medication regiment prescriptions, and alerting for patient well-being. In this capacity, sensors can be applied both in in-patient and out-patient care. Heterogeneous wireless access-based remote patient monitoring systems can be deployed to reach the patient everywhere, with multiple wireless [18].

It is possible to single out some additional functionalities of Healtcare-IoT described in the work [19],[20], such as cross-organization integration. Hospital Information Systems (HIS) should be extended with the patient-house, and can be integrated in a large part of the health system that can cover a community, city or country.

III. HEALTHCARE-IOT THROUGH ARCHITECTURAL LAYERS

Understanding of the IoT is possible through marking the common aspects and new technologies in architectural layers, in order to create complex intelligent solutions in the field of healthcare. New objects are becoming part of the digital process for the realization of remote smart services in the health system.

Known Research Projects which provide different aspects of IoT architecture are OpenIoT [21], IoT @ Work [22], iCore [23], SENSEI [24], PECES [25], SemSorGrid4Env [26], U2IoT [27].

On Fig.1 is presented integrated Healthcare system extended with IoT for Healthcare which consists of several layers: Business Layer, Sensor Layer, Network Layer, Service Layer, Application Layer.

Cloud computing for Healthcare-IoT can provide a stable and low cost Infrastructure as a Service (IaaS) that will support the sensors and actuators, such as Platform as a Service (PaaS) for accessing, processing and placing the big quantities of data.

Architectural layers of Healthcare-IoT enable the monitoring of data collection and transfer of vital information from a patient to a doctor, as well as the tracking of all the resources of the health system, data interoperability and integration with inherited hospital systems.

A. Business Layer

Business Layer should create a clear business model for new services in healthcare, give objective, functionalities, define business processes, roles and extract the technologies that best suit the required task. Some of the possible scenarios that can be modeled are the cases of remote in-home monitoring of chronically ill patients, response to emergencies, monitoring the recovery of patients after treatment, patients' response to treatment, methods of prevention, monitoring of hospital resources. It is necessary to create a new value chain so that existing services in an integrated traditional health system can expand to new services. Business Layer of IoT Healthcare introduces a new level of healthcare which can be called in-home or anywhere, anytime healthcare.

B. Sensor Layer

The lowest layer should provide a new source of information for healthcare. Numerous sensors could be implemented such as Blood Pressure Sensor (sphygmomanometer), Body Temperature Sensor, Glucometer Sensor, Sensor Electrocardiogram (ECG), Pulse and Oxygen in Blood Sensor (SPO2), Patient Position Sensor (Accelerometer), Airflow Sensor (Breathing), Galvanic Skin Response (GSR). Sensors, actuators, gateways and storage systems have the ability to feel or change their environment. Biomedical Signals should be created with the help of special devices attached to the patient's body or special carriers such as wireless body sensors and on / off-body networks technologies. In line with other multimedia information concentrated around the patient, most applications are based on the data collected from video cameras, microphones, movement and vibration sensors.

According to [5], there are several examples of the use of high-tech clothing as sensors. The LifeShirt System is the first non-invasive, continuous ambulatory monitoring system that can collect data as a cardiopulmonary function and other physiological patient parameters, and correlate them over time. The high tech vest uses optical fibers to detect bullet wounds and monitor the body vital signs during combat conditions. The patient monitoring finger ring sensor measures PPG signals, skin temperature, blood flow, blood constituent concentration and the pulse rate of the patient. The data are encoded for wireless transmission by mapping a numerical value associated with each datum to a pulse emitted after a delay of a specified duration, following a fiducial time. Multiple ring bands and sensor elements may be employed to determine threedimensional dynamic characteristics of arteries and tissues. These data may be transmitted wirelessly for further analysis. At this level, there is limited data processing and data selection to save resources. Data can be processed on different layers and software components provide information or allow the activity of the facilities. Processing large amounts of data can be done in the Cloud where the data can be placed and perform complex processing at the user's request [5].

C. Network Layer

Network Layer should provide a high performance of network infrastructure as a transport medium for IoT data. Healthcare-IoT services and applications make considerable demands such as speed transactional services which can be solved by using several networks with different technologies and protocols, and communication requirements for latency, throughput and security. Wireless technologies have the advantage in the realization of the concept of smart services anywhere and anytime. Unfortunately, there are no perfect wireless protocols (Satellite, Cellular, Wi-Fi, Bluetooth, Zigbee), which depend on the current solutions of heterogeneous health applications. Mobile health systems are developed on WiFi (Wireless Fidelity), GPRS (General Packet Radio Service) and 3G UMTS (Universal Mobile Telecommunications System) networking technologies. The current development of new communication technology improves some of the limitations of the existing wireless technologies and offers solutions such as E-Health via High-Speed Satellite Networks, E-Health via High-Speed Mobile Networks, HSPA (High Speed Packet Access), Personal Area Networks: on-body (wearable) and off-body networks [5].

D. Service Layer

Service Layer need to manage data and services of IoT. Data management is the ability to manage data information flow and provide information in the form of events or contextual data. Some events require later processing at certain periods of time, while other events, such as emergencies, require simultaneous processing. Various analytics tools are used to extract relevant information from a massive amount of raw data, which need to be processed at a much faster rate. In-memory analytics and streaming analytics are forms of analytics in real-time. The business and process rule engines should organize healthcare services through orchestration and choreography.

E. Application Layer

Applications are on the top of the architecture and need to export all the system's functionalities to the final user. Through the use of standard web service protocols and service composition technologies, applications can realize a perfect integration between distributed systems and applications [8]. There are various applications that can be involved in Health-IoT. Tele-medicine applications cover the areas of emergency healthcare, homecare, patient telemonitoring such as tele-cardiology, tele-radiology, tele-pathology, tele-dermatology, tele-ophthalmology, tele-psychiatry and tele-surgery, elderly family member monitoring, continuous patient monitoring, monitoring (smart) pills. These applications enable the provision of prompt and expert medical

services in underserved locations, like rural health centers, ambulances, ships, trains, airplanes as well as at homes.

The presented integrated Healthcare system should include new healthcare services in the traditional Healthcare system. Interoperability of devices and services from different suppliers, operational workflow should be solved. Security, privacy and trust must be enforced across the whole dimension of the Healtcare-IoT system. Healthcare-IoT solution should involve the technologies that best match the specified model.

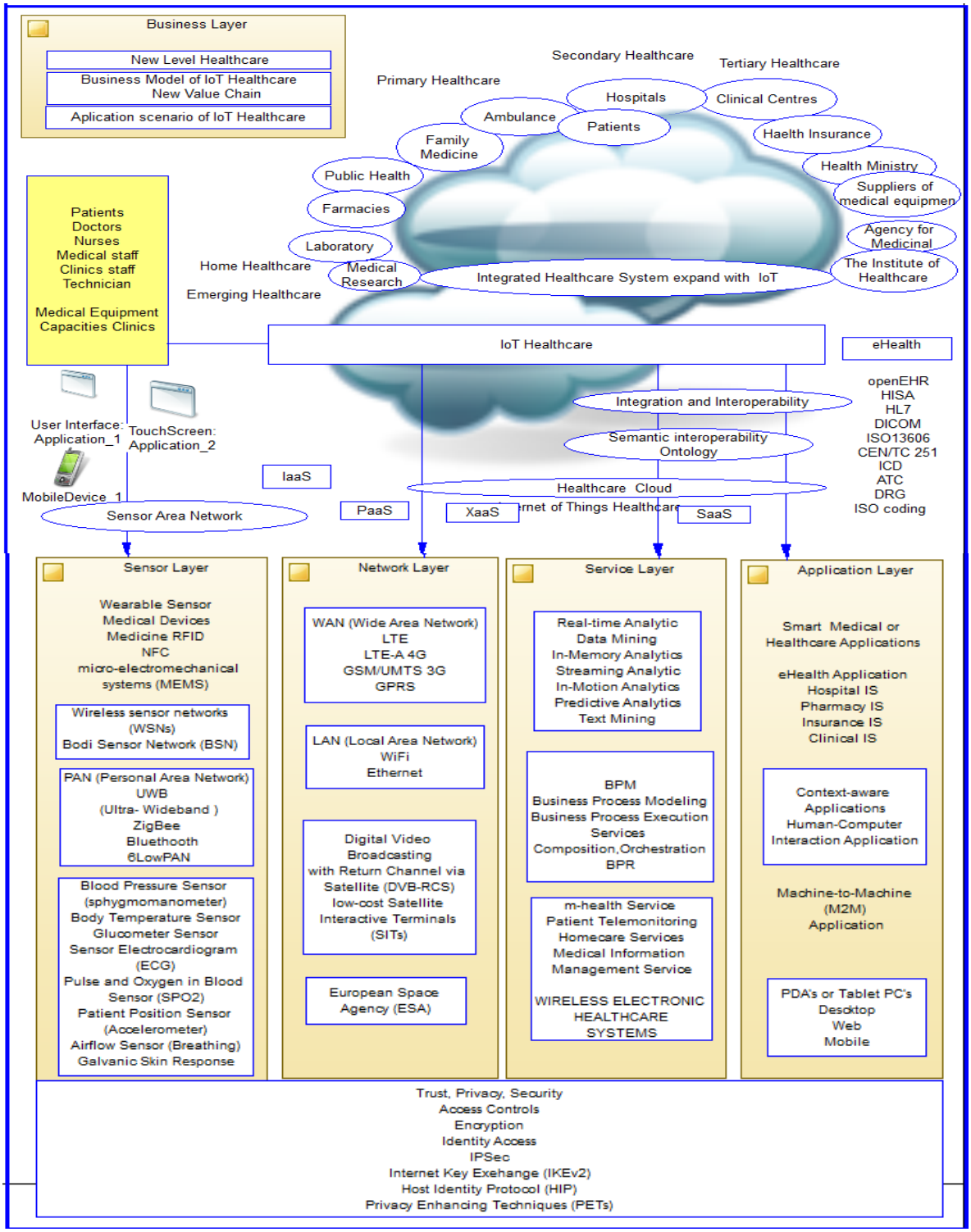


Figure 1. Integrated Healthcare system extended with IoT for Healthcare throught Architectural Layers

IV. TECHNOLOGIES FOR IOT AND COMMUNICATIONS OPPORTUNITIES

The Internet of Things requires the integration of modern technologies and communications solutions and is a result of synergetic activities in different fields of knowledge, such as telecommunications, informatics, electronics and social science.

Several technology trends will help shape IOT. Here are seven identified macro trends: the miniaturisation of devices, advances in RFID technologies, Internet Protocol version Six (IPv6), improvements in communication throughput and latency, real-time analytics, adoption of cloud technologies and security [12].

Miniaturisation of devices – The size (and cost) of electronic components that are needed to support capabilities such as sensing, tracking and control mechanisms, play a critical role in the widespread adoption of IOT for various industry applications. The progress in the semiconductor industry is spectacular. Due to the decreasing size and cost of technology components, organisations will see greater savings and opportunities in pursuing IOT in the next one to three years.

Radio Frequency Identification (RFID) is a technology of particular importance to IOT since one of the first industrial realisations of IOT, is in the use of RFID technology to track and monitor goods in the logistics and supply chain sector.

Internet Protocol version 6 (IPv6) is the next Internet addressing protocol that is used to replace IPv4. With IPv6, there are approximately 3.4×1038 (340 trillion trillion trillion) unique IPv6 addresses, allowing the Internet to continue to grow and innovate.

Increasing communication throughput and lower latency enhance the infrastructure for the support of data capability and improve network throughput with the addition of General Packet Radio Service (GPRS), Global System for Mobile (GSM), Enhanced Data rates for GSM Evolution (EDGE), Universal Mobile Telecommunications System (UMTS) with High Speed Packet Access (HSPA), Wideband Code Division Multiple Access (WCDMA), Long Term Evolution (LTE). Low latency makes it possible for IOT applications to query or receive quicker updates from sensor devices.

Real-time Analytics – In-memory processing is a form of analytics where detailed data is loaded into the system memory from a variety of data sources. New data are analysed and stored in the system memory to improve the relevance of the analytics content and to augment the speed in decision making. Another form of real-time analytics, such as streaming analytics, uses complex algorithms to instantaneously process streams of event data it receives from one or more sources.

Cloud Computing - Cloud Infrastructure as a Service (IaaS) uses hardware such as sensors and actuators, which can be made available to consumers as cloud resources. Cloud Platform as a Service (PaaS) can provide a platform from which to access IOT data and on which custom IOT applications (or host-acquired IOT applications) can be developed. Cloud Software as a Service (SaaS) can be provided on top of the PaaS solutions to offer the provider's own SaaS platform for specific IOT domains.

Security and Privacy – IPv6 contains IPSec, access control, connectionless integrity, data origin authentication, protection against replays (a form of partial sequence integrity), confidentiality (encryption), and limited traffic flow confidentiality. Other IP-based security solutions such as Internet Key Exchange (IKEv2) and Host Identity Protocol (HIP) are also used to perform authenticated key exchanges over IPSec protocol for secure payload delivery. For technical implementations, there are Privacy Enhancing Techniques (PETs) such as anonymisation and obfuscation to de-sensitize personal data. Extensible Authentication Protocol (EAP) is an authentication framework used to support multiple authentication methods. Protocol for carrying Authentication for Network Access (PANA) forms the network-layer transport for EAP.

The impact caused by the IoT on human life will be as huge as the one that the internet has caused in the past decades, so the IoT is recognized as "the next of internet". A part of the included technologies are sensors and actuators, Wireless Sensor Network (WSN), Intelligent and Interactive Packaging (I2Pack), real-time embedded system, MicroElectroMechanical Systems (MEMS), mobile internet access, cloud computing, Radio Frequency IDentification (RFID), Machine-to-Machine (M2M) communication, human machine interaction (HMI), middleware, Service Oriented Architecture (SOA), Enterprise Information System (EIS), data mining, etc. With various descriptions from various viewpoints, the IoT has become the new paradigm of the evolution of information and communication technology [20].

The combination of sensors, RFID, NFC (near field communication), Bluetooth, ZigBee, 6LoWPAN, WirelessHART, ISA100, Wi-Fi will enable significantly improved measurement methods and monitoring of vital functions (temperature, blood pressure, heart rate, cholesterol, level of glucose in the blood) [5].

Key challenges for IoT according to [28] are: technical, security, privacy, as well as trust, societal, business challenges.

CONCLUSIONS

This work has presented the possibilities of expanding the complex healthcare system with new services Healthcare-IoT as a process of creating new values for the health system.

The Fig. 1 has marked the common aspects and new technologies through architectural layers, which enabled the consideration of the application of intelligent healthcare solutions. Business layer should create a clear model of healthcare.

The new models of health care can be called healthcare anywhere and anytime or home healthcare. The existing services in the integrated healthcare system need to be expanded with new services.

This paper has taken into consideration some of the grouped Healthcare-IoT functionalities from the literature. Some of the presented benefits of Healthcare-IoT are: monitoring of patients, staff and equipment of hospitals in order to provide better workflows for hospitals.

Identification and authentication can help to reduce errors, as well as procedures and treatment of patients with the wrong medication. Real-time information about the

patient's health indicators, collected from the sensors, may be available to doctors anywhere and anytime.

This paper has also presented the technologies which are under development and which should enable a simple and cheap implementation of IoT solutions. Wireless and sensor technologies promise a reliable and safe collection of health indicators about the patient. There are still many open issues and research challenges to address such as the wide distribution of the objects, dependable communications to work with a weak radio signal, propagation through the human body, efficient data compression, cheaper chips, data rates, real-time processing. Depending on the choice of the business objectives for health care and precisely defined business processes in Business Layer, it is necessary to choose suitable technologies that will enable a successful implementation of IoTs.

This paper has provided a holistic approach to Healthcare-IoT that should remove the gap between individual technology solutions and business innovation in healthcare as an integrated Healthcare system expanded with IoT functionalities.

Further research will be in the field of application of technologies in the form of specific solutions for connecting a particular Healthcare-IoT solution with the existing healthcare system.

REFERENCES

[1] D. Giusto, A. Iera, G. Morabito, L. Atzori (Eds.), The Internet of Things, Springer 2010, ISBN: 978-1-4419-1673-0.

[2] Ashton, K., (2009). I could be wrong, but I'm fairly sure the phrase "Internet of Things" started life as the title of a presentation I made at Procter & Gamble (P&G) in 1999.

[3] Brock, D. L., (2001). The Electronic Product Code: A Naming Scheme for Physical Objects. White Paper of MIT Auto-ID Center, Jan 2010, MIT-AUTOID-WH-002.

[4] ITU Internet Reports 2005: (2005). The Internet of Things-Executive Summary [online]. Available from:www.itu.int/osg/spu/publications/internetofthings [Accessed 10.2014]

[5] D. Vouyioukas, I. Maglogiannis, "Communication Issues in Pervasive Healthcare Systems and Applications Pervasive and Smart Technologies for Healthcare: Ubiquitous Methodologies and Tools", IGI Press Pages 197-227 2010.

[6] Tam Harbert. FCC Gives Medical Body Area Networks Clean Bill of Health. [Online] Available from: http://spectrum.ieee.org/tech-talk/biomedical/devices/fcc-gives-medical-body-area-networks-clean-bill-of-health [Accessed 2014].

[7] Breynaert, D. (2005). 2Way-Sat: A DVB-RCS Satellite Access Network. Technical Report, Newtec Cy N.V. (NTC), Belgium.

[8] Limburg, M.; Gemert-Pijnen, J. E.; Nijland, N.; Ossebaard, H. C.; Hendrix1, R.; Seydel,E., (2011). Why Business Modeling is Crucial in the Development of eHealth Technologies. Journal of Medical Internet Research, 13(4), e124,doi:10.2196/jmir.1674

[9] World Health Organization (2011). mhealth: New Horizon for Health Through Mobile Technologies. Global Observatory for e-Health Services, vol. 3. Geneva, Switzerland: WHO.

[10] Pervasive or Ubiquitous Healthcare?, B. Arnrich1; O. Mayora2; J. Bardram3, Methods Inf Med 2010; 49: 65–66.

[11] Pawar, P.; Jones, V.; Beijnum, B. F. V.; Hermens, H., 2012. A framework for the comparison of mobile patient monitoring systems. *Journal of Biomedical Informatics,* 45(3), 544-556.

[12] InternetOfThings [Online]. https://www.ida.gov.sg/~/media/Files/Infocomm%20Landscape/Technology/TechnologyRoadmap/InternetOfThings.pdf [Accessed 06. 2014]

[13] Koop, C.E.; Mosher, R.; Kun, L.; Geiling, J.; Grigg, E.; Long, S.; Macedonia, C.;Merrell, R.; Satava, R.; Rosen, J. (2008). Future delivery of health care: Cybercare.IEEE Engineering in Medicine and Biology Magazine, 27(6), 29-38.

[14] A. Jara, M. Zamora, and A. Skarmeta, "Knowledge acquisition and management architecture for mobile and personal health environments based on the internet of things," in Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on, 2012, pp. 1811–1818.

[15] Cluster of European Research Projects on the Internet of Tings, CERPIoT. 2009. Internet of Tings: Strategic research roadmap

[16] A.M. Vilamovska, E. Hattziandreu, R. Schindler, C. Van Oranje, H. De Vries, J. Krapelse, RFID Application in Healthcare – Scoping and Identifying Areas for RFI D Deployment in Healthcare Delivery, RAND Europe, February 2009.

[17] Atzori, L., Iera, A., Morabito, G. (2010). The Internet of Things: A survey, (Jurnal Elsevier)Computer Networks, 54(15), 2787-2805

[18] D. Niyato, E. Hossain, S. Camorlinga, Remote patient monitoring service using heterogeneous wireless access networks: architecture and optimization, IEEE Journal on Selected Areas in Communications 27 (4) (2009) 412–423.

[19] Serbanati, L. et al, 2011. Steps towards a digital health ecosystem. In Journal Biomedical informatics, doi:10.1016/j.jbi.2011.02.011

[20] Z. PANG, Technologies and Architectures of the Internet-of-Things (IoT) for Health and Well-being, Doctoral Thesis in Electronic and Computer Systems KTH – Royal Institute of Technology Stockholm,Sweden, January 2013.

[21] O. Consortium. Open Source Solution for the Internet of Things intothe Cloud - OpenIoT. [Online]. Available: http://www.openiot.eu/ [Accessed 10.2014]

[22] I. Consortium. Internet of Things at Work - IoT@Work. [Online]. Available: http://www.iot-at-work.eu/ [Accessed 10.2014]

[23] iCore Consortium. Empowering IoT through Cognitive Technologies -iCore. [Online]. Available: http://www.iot-icore.eu/ [Accessed 11.2014]

[24] SENSEI: An Architecture for the Real World Internet, SISS 2010, http://www.ieee-scc.org/2010/web_pages /SISS/presentation/keynote.pdf [Accessed 11.2014]

[25] Pervasive computing in embedded systems. FP7. http://www.ictpeces.eu [Accessed 05. 2014]

[26] Semantic sensor grids for rapid application development for environmental management. FP7. http://www.sem sorgrid4env.eu/ [Accessed 10. 2014]

[27] Huansheng Ning, Unit and Ubiquitous Internet of Things, Book Published: April 4, 2013 by CRC Press

[28] Finnish Strategic Centre for Science, Technology, and Innovation:For Information and Communications (ICT) Services, businesses, and technologies, Version 1.0 1st September 2011.

Limitations of Smartphone MEMS for motion analysis

Anton Umek and Anton Kos
Faculty of Electrical Engineering, University of Ljubljana, Slovenia
anton.umek@fe.uni-lj.si , anton.kos@fe.uni-lj.si

***Abstract*—The paper presents the limitations of smartphone motion sensors and their suitability for simple biofeedback applications with motion detection. We have measured the smartphone accelerometer and gyroscope biases, identified the main causes for short-term and long-term bias variations, and quantified their precision. Under certain conditions the existing smartphones with their built-in inertial sensors are suitable for use in real-time biofeedback applications.**

I. INTRODUCTION

Smartphones are readily available, wide-spread technology. According to [1] in many countries the penetration of smartphones has exceeded 50% in 2014. Practically all new smartphones available today are equipped with MEMS (Microelectromechanical systems) sensors, which are mostly low cost sensors and consequently of relatively low quality. Their use for motion tracking is limited due to their low quality which is expressed through the imprecision of sensor signals. MEMS accelerometer and gyroscope quality is mostly degraded by bias, noise and nonlinearity. Several calibration methods can improve the sensors precision, which is mostly affected by their biases . Most basic calibration method include sensor bias measurement and compensation. For an application developer the exact specifications of the inertial sensors integrated into smartphones are not always well-known. There are several different possible reasons for that, for example: specifications are not available, application runs on different hardware (android smartphones), or sensor readings are preprocessed inside the smartphone's operating system. Even if the inertial sensor specifications are known, it is reasonable that an application using them has functionality for the measurement of their properties. Measured values give us the ability to act in cases when the sensor precision is outside the bounds required by the application.

Our first goal was to explore the suitability of inexpensive inertial sensors integrated into today's smartphones for the development of a real-time motion biofeedback system. In a *biofeedback* system application user has attached sensors for measuring body functions and parameters (*bio*). Sensors signals are transferred to signal processing device and results are communicated back to the person (*feedback*) through one of the human senses (i.e. sight, hearing, touch) [2]. The person tries to act on received information to change the body motion in the desired way. As an example of a real-time biofeedback system we designed an application that helps users to correct specific golf swing errors [3]. The application uses the inertial sensors integrated into the smartphone, which is attached to the head of the golf player. With the appropriate attachment of the inertial sensor to the cap of the golf player, we can achieve very good repeatability of detection of different 3D head movements.

II. BIAS MEASUREMENTS

In our experiments we used iPhone 4 smartphones. As identified by Chipworks [4] and [5], the iPhone 4 embedded 3D gyroscope and 3D accelerometer integrated circuits are manufactured by STMicroelectronics. Main iPhone 4 sensor parameters, acquired from the manufacturer's data sheets [6] - [8] are listed in Table 1.

TABLE I.
THE MAIN PARAMETERS OF THE IPHONE 4 MEMS [6]-[8].

Parameter	3D accelerometer LIS331DLH	3D gyroscope L3G4200D
Range	±2 g	±2000 deg/s
Sensitivity	1±0.1 mg/dig	70 mdeg/s/dig
Bias error	±20 mg	±8 deg/s

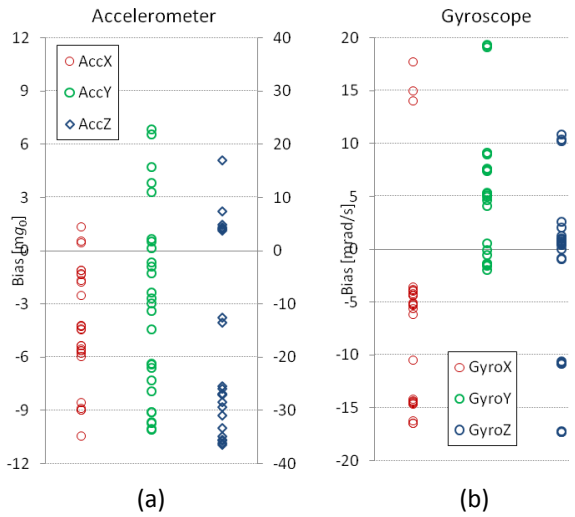


Figure 1. 3D accelerometer (a) and 3D gyroscope (b) biases gained from multiple measurements on several smartphones. Biases are calculated by averaging $N = 600$ sensor signal samples at sampling frequency $f_s = 60$ Hz, the corresponding averaging time is therefore $\tau = 10$ s. Accelerometer biases presented in mg_0 have different dynamic ranges and are presented on two separate scales; the left hand side scale is valid for X and Y axes, and the right hand side scale for the Z axis. Gyroscope biases, presented in mrad/s for all three axes, have similar dynamic ranges and are presented on the same scale.

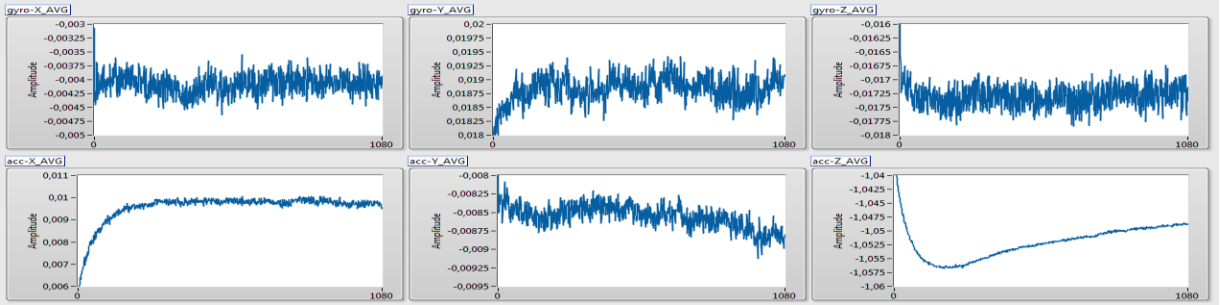


Figure 2. Measured bias values in the 3 hour temperature stress test. Each bias value is gained by averaging 600 sensor signal samples in 10 second intervals. Therefore 1080 bias values shown in graphs are gained in the 3 hours long test. Temperature changes induce noticeable bias drift of accelerometers and gyroscopes.

The precision of smartphone sensors is mostly affected by their biases, which induce errors in the derived angular and spatial position [9], [10]. There are other factors that influence the sensor precision. For example, a simple and computationally efficient calibration procedure for sensor axes misalignment is presented in [11].

For the estimation of the bias value ranges we have conducted a series of measurements on a number of iPhone 4 smartphones in different time intervals. Bias measurements were carried out with the specially developed application running in LabVIEW and using the specially designed and constructed casing allowing us to orientate the smartphone in any of the principal positions in a simple way. The aim of this measurement results is to present the bias variations on different smartphone devices of the same type (iPhone4).

The measurements were performed in twelve positions to eliminate the influence of the gravity. The application receives sensor data from the smartphone over the local wireless network and calculates a range of statistical parameters of smartphone inertial sensor signals. Bias measurement results for six iPhone 4 smartphones are shown in Fig. 1. Measurements for all three accelerometer and gyroscope axes are averaged over the time interval τ = 10 s. The gained gyroscope biases are within the range of ΔG_0 = ±20 mrad/s = ±1.15 deg/s. The gained accelerometer biases are within the range ΔA_0 = ±12 mg$_0$ for the X and Y axes, and ΔA_0 = ±40 mg$_0$ for the Z axis.

While the differences between devices are the effect of variations in physical properties of MEMS [8], the differences in successive measurements of the same device are caused by various inertial sensor instabilities, most probably because of slightly different internal phone temperatures between measurements. To test our assumptions, we conducted a simple temperature stress test. First, we have cooled the switched-off smartphone down to 8°C. After switching the phone on and putting it in the place with the room temperature of 22°C, we have measured biases in time periods of several hours. The bias drifts, shown in Figure 2, exhibit the strongest temperature dependence in the first 30 minutes of the test. During this time interval the temperature changes induce bias drifts that can be much larger than the bias variations caused by other factors of sensor instabilities, including noise. In the measurement of a particular smartphone, shown in Fig. 2, the latter is especially evident in the bias drift of the accelerometer axes X and Z. The results of stress test measurements on other smartphones show that strong temperature induced bias drifts are expressed also on other exes. Temperature stress tests show that the change in temperature, as it was anticipated, causes large bias variations, especially for accelerometers. Stable temperature conditions are reached after 60 minutes.

Bias variations measured in constant room temperature conditions are much smaller. The same measurements were carried out in the 60 minutes period on six different smartphones; each bias value is gained by averaging the 600 sensor signal samples in 10 second intervals. Accelerometer bias variations due to noise are in the range of 0.35 to 0.5 mg$_0$. Gyroscope bias variations due to noise are in the range of 0.7 to 1.0 mrad/s. Larger differences between smartphones are expressed in bias drifts. Gyroscope bias variations are primarily the result of the sensor white noise. In most cases the measured bias drift is comparable to the averaged bias noise. Less often the bias drift is higher than the averaged bias noise (τ = 10 s). In the worst case, the measured gyroscope drifts stay below 2 mrad/s per hour and the measured accelerometer bias drifts do not exceed 4 mg$_0$ per hour.

A. Allan variance measurements

Bias variations are caused by various random processes in the operation of the sensor. The Allan variance method helps us to determine the characteristics of the underlying random processes and noise models.

Biases are measured by averaging a finite sequence of samples when the device is in the standstill position. Bias approximations $y[m]$ are calculated by averaging the sensor signal samples $x[n]$:

$$y[m] = \frac{1}{N}\sum_{n=0}^{N-1} x[n + m \cdot N] \qquad (1)$$

Allan variance $\sigma_A^2(N)$ is a measure of variations of mean values $y[m]$ of the consecutive blocks of N signal samples $x[n]$ [12]:

$$\sigma_A^2[N] = \frac{1}{2}\overline{(y[m] - y[m-1])^2} \qquad (2)$$

The approximation of the variance is calculated from the finite number of mean values $y[m]$:

$$\sigma_A^2[N] \approx \frac{1}{2\cdot(M-1)}\sum_{m=1}^{M-1}(y[m] - y[m-1])^2 \qquad (3)$$

Gathering the data for Allan variance calculation $\sigma_A^2(\tau = N T_s)$ requires long measurement times τ and also a large number of signal samples $x[i]$. The upper averaging time was set to τ_{max}=1000 s. Allan variance

measurements are performed simultaneously for five different averaging times $\tau = \{0.1, 1, 10, 100, 1000\}$ s. The maximal averaging time requires $N_{max} = 60000$ signal samples. For a statistically relevant measurement the minimum number of averaging episodes must be $M_{min} = 10$. Therefore the total measurement time $T_0 = 10000$ s.

Depending on the nature of the random process, the bias noise has different power spectrum shapes. Noises with different spectrum power density profiles appear in the Allan variance plot with different slopes [12]. In such situation it is possible to identify the model of the underlying random process from the Allan deviation $\sigma_A(t)$ log-log plot, where different noises appear in different regions of τ. Allan variance measurements of the 3D accelerometer and 3D gyroscope for a single smartphone are shown in Fig. 3. As shown in Fig. 3(a) the accelerometer Allan deviation $\sigma_A(t)$ is following the slope of the bias white noise model at short averaging times $\tau \leq 10$ s. Accelerometer *velocity random walk* constant (VRW) can be determined from the Allan deviation plot at $\tau = 1$ s; VRW = $\sigma_A(\tau = 1$ s). Model parameters for all three axes are given inside the shaded rectangle in Fig. 3(a).

As shown in Fig. 3(b) the gyroscope Allan deviation $\sigma_A(t)$ is following the slope of the bias white noise model at short averaging times $\tau < 100$ s. Gyroscope *angle random walk* constant (ARW) can be determined from the Allan deviation plot at $\tau = 1$ s; ARW = $\sigma_A(\tau = 1$ s). Model parameters for all three axes are given inside the shaded rectangle in Fig. 3(b). At longer averaging times, where the averaging filter decreases the power of high frequency white noise, slow bias fluctuations with accented low frequency spectrum becomes the dominant error source for accelerometers and gyroscopes. Log time resolution measurements at $\tau = \{0.1, 1, 10, 100, 1000\}$ s in Fig. 3 allow only the determination of the minimal bias instabilities (BI), which are expressed at $\tau = 100$ s for the accelerometer and $\tau = 1000$ s for the gyroscope.

Measurements were conducted under the conditions of the stable smartphone operation: constant room temperature, absence of vibrations, and low and constant power dissipation of the smartphone. The same measurements were performed on six different iPhone4 devices. We have noticed only minor differences in noise models. The calculated average sensor parameters are: VRW = 0.26 $mg_0/\sqrt{Hz}$, ARW = 26 mdeg/s/$\sqrt{Hz}$. More noticeable are the differences in the bias instability, which are to our belief primarily the effect of different temperature sensibilities of the smartphones.

On the basis of the average sensor model we can determine averaging times τ for bias measurement that is used as offset value at sensor bias compensation. We conclude that the reasonable averaging times for bias measurements are between 10 s and 100 s.

B. Bias estimation error

The measurements of accelerometer and gyroscope biases on several smartphones indicate that some particular applications with moderate demands for accuracy and short time analysis could use even uncompensated sensor data. From Fig. 1 we see that gyroscope biases are less than 1 deg/s, and accelerometer biases are below 40 mg_0. Uncompensated bias values are inducing *bias error (1)* from Fig. 4(b). For the *uncompensated gyroscope* this means that in a short time analysis, for instance in a 3 seconds interval, the bias induces angular error of less than 3 degrees. Relatively high accelerometer biases restrict the use of *uncompensated accelerometers* for position tracking, but for the calculation of the tilt angle between a smartphone's axis and the gravitation vector, the error is less than 2 degrees.

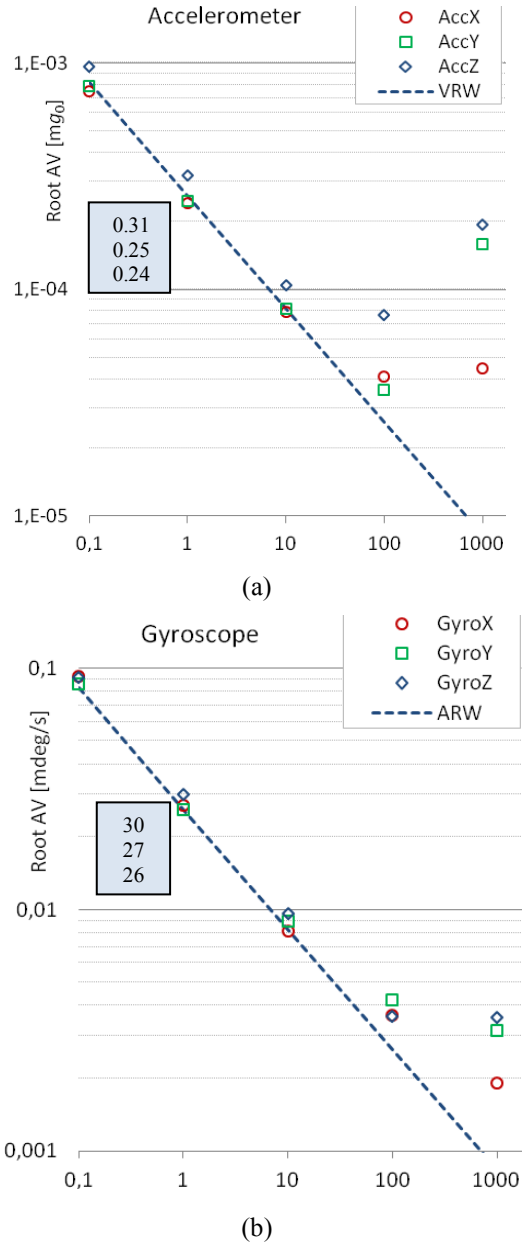


Figure 3. Allan variance measurements for all three axes of the accelerometer and gyroscope of the single smartphone. (a) Accelerometer results conform to the VRW model at short averaging times. (b) Gyroscope results conform to the ARW model at short averaging times.

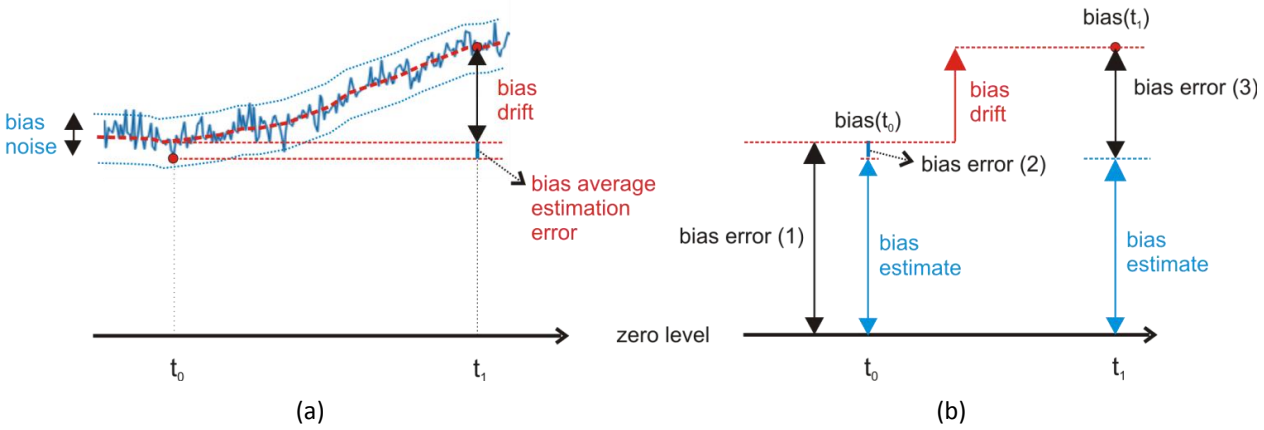


Figure 4. Inertial sensor bias variations and their effects. The measured bias changes with time (blue line), in the short time primarily because of bias noise (ARW, VRW), in the long term because of other influences (red dashed line). The instantaneously acquired bias estimation value at the time t_0 (red dot) can differ from real bias for the estimation error (the difference between the parallel red dotted lines). Bias drift is the change in bias value between times t_0 and t_1. (b) Bias error is the difference between the *zero level* and the real bias value. Without compensation we experience *bias error* (1), with the compensation at time t_0 we decrease the bias error for the measured bias estimate to get *bias error* (2). By the time t_1 the bias drift causes the error to grow to the value of *bias error* (3).

The precision of the accelerometer and the gyroscope is considerably better after bias compensation. The compensated bias values at time t_0 are inducing bias error (2) from Fig. 4. Gyroscope bias compensation can be performed in a smartphone's standstill position. Based on Allan variance results from previous section, the sufficient averaging time $\tau = 10$ s. This gives us a good compromise of the noise elimination and bias instability influences. Accelerometer bias compensation requires several successive measurements in precise rotations around the three axes. Consequently the compensation of the accelerometer takes much more time and effort.

The calculated variation of the mean bias values is $\Delta G_0 = 25$ mdeg/s for the gyroscope and $\Delta A_0 = 0.25$ mg$_0$ for the accelerometer. That means that 3 seconds after the bias compensation the maximal angular drift is 0.075 degree and maximal position drift is 1.1 cm. According to Fig. 4., the above values are valid only shortly after the bias compensation at time t_0. If we perform the same analysis at time t_1 the expected angular and velocity drift would be higher because of bias drift that would result in *bias error (3)*. If a new bias error is not acceptable for the application, another bias compensation procedure should be performed at time t_1.

III. BIAS COMPENSATION OPTIONS

Inertial sensors bias variations in the form of noise and drift could be the limiting factor for their usability in different types of applications. In the biofeedback applications, where we generally use inertial sensors to measure movement patterns, large biases are a limiting factor. The precision of the sensor readings can be improved to a certain extent by bias compensation, but we have to bear in mind that bias errors can never be fully eliminated as it is evident from Fig. 4. With regard to each individual application and its demands for sensor precision, we must choose the correct bias compensation strategy:

- *One-time, single bias compensation* has a time-limited effect. Therefore a one-time compensation is suitable only for applications that operate in a stable environment. Periodic bias measurements on the same smartphone, in long time intervals over several weeks, showed that estimated gyroscope biases vary for less than 30 mdeg/s and the estimated accelerometer biases vary for less than 1 mg$_0$.
- *Periodic, repetitive bias compensation* can be performed in regular time intervals or on as needed basis. For instance, the bias compensation is needed at every significant change of the inertial sensor temperature. This temperature change can be caused by the change of the ambient air temperature or by the change of the inner temperature of the smartphone. In our experience, confirmed by the measuring results, the change in the inner temperature caused by running applications causing the heating of other hardware and integrated circuits of the smartphone, has greater and more instant effect than the change of the ambient air temperature.

To achieve different levels of movement detection accuracy the following strategies are possible:

(a) The application uses *uncompensated sensor data*. In this case the bias error corresponds to the *bias error 1)* in Fig. 4. The derived angle and position errors, as measured in previous section are 1 deg/s and 19 cm after the first second respectively. This compensation scenario could be applicable to short time movements, up to a few seconds long, if the application does not use accelerometers for position calculation and does not demand high angular precision.

As an example, signal analysis in our golf swing biofeedback application takes less than 3 seconds. Predicted gyroscope angular error in such short time

interval is less than 3 degree. Static position angle (tilt) is calculated from accelerometer data and maximal tilt error in vertical position is less than 2 degrees.

(b) Before use, the application performs a *one-time bias measurement and compensation* of the accelerometer and the gyroscope. In this case the bias error corresponds to the *bias error (2)* in Fig. 4. The effect of one-time compensation is satisfactory if the operating conditions do not differ much from the conditions at which the compensation was performed. In such cases biases change in a very limited value range. The derived angle and position errors, as measured in previous section, are 25 mdeg/s and 1.2 mm after the first second respectively. This compensation scenario could be applicable to short time movements, up to a few seconds long, even for the applications demanding high precision or for the medium-time movements, up to a few tens of seconds long for less demanding applications. When the operating conditions change considerably, new bias measurement must be carried out and if needed compensated.

One-time accelerometer and gyroscope bias compensation have significantly improved angle precision in our golf swing real-time biofeedback application. For a short time after the bias compensation both angular errors stay negligible small: vertical positioning (tilt) error is less than 15 mili-degrees and gyroscope rotation angle error in three seconds long time interval is less than 0.08 degree. Biases drifts and after one hour enlarges both angular positioning errors: vertical positioning angle error is less than 0.25 degree and predicted gyroscope rotation angle error is less than 0.35 degree.

(c) The application *constantly measures and compensates biases*. Compensation of gyroscopes is possible during the application use, while the compensation of accelerometers require temporary interruption of application use. At every detected opportunity, when the device is in standstill long enough, the gyroscope biases are measured, evaluated and if needed compensated. The measurement times required for this compensation scenario could be between 10 s and 100 s. After a longer time period without compensation, the application may notify the user that the accuracy of the operation might be compromised and that a new bias compensation is required.

IV. CONCLUSION

Our study shows that under certain conditions the existing smartphones with their built-in inertial sensors are suitable for use in real time biofeedback applications. We have studied accelerometer and gyroscope biases, identified the main causes for short-term and long-term bias variations, and quantified their precision. The uncompensated smartphone sensor biases are in the range of ±40 mg_0 for accelerometer readings and in the range of ±1 deg/s for the gyroscope readings. These values restrict their use in biofeedback systems to short analysis window, up to a few seconds long. Bias compensations can significantly broaden the range of sensor usability, both in the prolonged analysis times and in higher precision required. The compensation using averaging times between 10 s and 100 s primarily eliminates the sensor's white noise. It reduces the bias by the factor of magnitude. The long-term bias variations caused by temperature changes may remain a problem.

REFERENCES

[1] Global smartphone-penetration 2014. Available online: https://ondeviceresearch.com/blog/global-smartphone-penetration-2014

[2] Biofeedback definition. Available online: http://www.mayoclinic.org/tests-procedures/biofeedback/basics/definition/prc-20020004

[3] Anton Umek, Sašo Tomažič, and Anton Kos, "*Autonomous Wearable Personal Training System with Real-Time Biofeedback and Gesture User Interface*", Proceedings of the 2014 International Conference on Identification, Information and Knowledge in the Internet of Things, pages 122-125, October 2014, Beijing, China

[4] Motion sensing in the iPhone 4: MEMS accelerometer. *Available online*: http://www.memsjournal.com/2010/12/motion-sensing-in-the-iphone-4-mems-accelerometer.html

[5] Motion sensing in the iPhone 4: MEMS gyroscope. *Available online*: http://www.memsjournal.com/2011/01/motion-sensing-in-the-iphone-4-mems-gyroscope.html

[6] ST Microelectronics. MEMS digital output motion sensor ultra low-power high performance 3-axes "nano" accelerometer, LIS331DLH Specifications. *ST Microelectronics*, July 2009

[7] ST Microelectronics. MEMS motion sensor: ultra-stable three-axis digital output gyroscope, L3G4200D Specifications. *ST Microelectronics*, December 2010

[8] ST Microelectronics. Everything about STMicroelectronics' 3-axis digital MEMS gyroscopes, TA0343 Technical article. *ST Microelectronics*, July 2011

[9] Mohinder Grewal; Angus Andrews. How good is your gyro. *IEEE Control Systems Magazine*, February 2010

[10] Harvey Weinberg. Gyro mechanical performance: the most important parameter. Technical article MS-2158, *Analog devices*, 2011

[11] Sara Stančin, Sašo Tomažič. Time- and computation-efficient calibration of MEMS 3D accelerometers and gyroscopes. Sensors, ISSN 1424-8220, Aug. 2014, vol. 14, no. 8, pages 14885-14915

[12] Naser El-Sheimy; Haiying Hou; Xiaoji Niu. Analysis and Modeling of Inertial Sensors Using Allan Variance. *IEEE Transactions on Instrumentation and Measurement*, Volume 57, 140-149

Segmentation and Three-Dimensional Visualization of Brain Tumor and Possibility of Mapping Such Algorithms on High Performance Reconfigurable Computers

Tijana Šušteršič [1], Nikola Mijailović [1], Ivan Milanković [1, 2], Nenad Filipović [1], Aleksandar Peulić[1]
[1] Faculty of Engineering, University of Kragujevac, Serbia
[2] Research and Development Center for Bioengineering, BioIRC, Kragujevac, Serbia
tijana.sustersic93@gmail.com, nmijailovic@kg.ac.rs, ivan.milankovic@kg.ac.rs, fica@kg.ac.rs, aleksandar.peulic@kg.ac.rs

***Abstract*—Brain tumor detection and visualization have important role in the process of the surgical operation, postoperative recovery and disease progression. In this paper methodology and algorithms for brain tumor extraction from the surrounding tissue and boosting the intensity of the tumor region is presented. The methodology is based on the fact that tumor tissue has different average gray value than healthy soft brain tissue. The process of segmentation with depth first search procedure is employed. Also, the possibility of mapping such algorithm on High Performance Reconfigurable Computers (HPRCs) is described. The input data used in this study are the set of the multi slice CT images. The obtained results and 3D reconstructed object can be new paradigm in the surgical assistance tools and brain tumor treatment.**

I. INTRODUCTION

The adult body normally forms new cells only when they are needed to replace old or damaged ones. In infants and children, new cells are formed during development in addition to those needed for repair. A tumor develops if normal or abnormal cells multiply when they are not needed. A tumor is a mass of tissue that grows out of control of the normal forces that regulate growth [1]. In addition, a brain tumor is classified as a mass of unnecessary cells growing in the brain or central spine canal. These and other tumors in the body are due to changes (mutations) in specific genes of normal cells. Mutations of some genes become active (prooncogenes become oncogenes) and the others are inactivated (tumor suppressor genes). This process happens all the time in the organism in many cells, but the immune system recognizes the cells and neutralize them. Sometimes these cells are not neutralized and, if they continue to multiply, they form a tumor [2].

There are two basic kinds of brain tumors - benign and cancerous (malignant) tumors. The difference between malignant and benign tumor is the aggressiveness in growth as well as the fact that malignant form metastasis and spread into the environment, infiltrating the surrounding tissue, and benign tumors do not give metastases to other organs and do not infiltrate into the surrounding healthy tissue, but are repressed, or grow expansively. Yet the difference between benign and malignant brain tumors is less distinctive than in other tumor systems, because brain tumors can lead to severe neurological disorder and death. In brain tumor, there is an increase in pressure within the cranial cavity due to the growth of the tumor mass, which mostly affects white brain mass, or due to obstructions in the normal circulation of brain fluid. When a tumor reaches a considerable size, it forms ambient pressure on the brain, leading to its swelling; blood flow is in disorder which eventually can lead to bleeding. The mechanisms by which tumor damages the tissues are numerous and, between themselves, crisscrossed [3].

Tumors vary in shape, size, location, and internal texture, and hence tumor segmentation is known to be a very challenging problem [4]. In practice, radiation oncologists spend a substantial portion of their time performing the segmentation task manually, using one of the available visualization and segmentation tools [5]. Therefore, the driving problem discussed in this paper is the segmentation of 3D brain tumors from computational tomography image data.

II. BACKGROUND

The extraction of the anatomical parts of the brain is a key problem in medical image analysis. Firstly, methods for recording medical images are developed using Computational Tomography – *CT*, Magnetic Resonance Imaging – *MRI,* Positron emission tomography – *PET* etc. that will accurately show the condition of the patient.

Image segmentation is an important and challenging factor in the medical image segmentation. Generally speaking, segmentation involves the separation of typical parts of an image for their further detailed analysis and discussion [6]. The Segmentation of an image entails the division or separation of the image into regions of similar attribute. The ultimate aim in a large number of image processing applications is to extract important features from the image data, from which a description, interpretation, or understanding of the scene can be provided by the machine. The digital image processing community has developed several segmentation methods whereas four methods for detecting tumors, edema and necrotic tissues are the most common [7]: 1) amplitude thresholding, 2) texture segmentation 3) template

matching, and 4) region-growing segmentation. These types of algorithms are based on the separation of characteristic parts using threshold, texture, or comparisons with the template and are used for dividing the brain images into three categories (a) Pixel based (b) Region or Texture Based (c) Structural based [2]. In this paper, more attention is given to pixel based algorithms due to their simplicity in functioning.

Computed Tomography (CT) scan is a safe, noninvasive test that uses an X-ray beam and a computer to make 2-dimensional images of the brain. Similar to an MRI, it views the brain in slices, layer by layer, taking a picture of each slice. A dye (contrast agent) may be injected into bloodstream.

Since the concepts of visualization of patient's state through CT images and 3D model of manipulable computer data are connected and start to overlap with surgical intervention, it is important to run a set of programmes in order to render 3D model out of a series of CT images. This can further lead to localization of a tumor and determining the optimal course for approaching a tumor. The surgeon can easily manipulate the 3D model in order to have an insight of the patient from any angle and depth. This method can help the surgeon to better understand the state of the patient and to establish a more precise diagnosis. Furthermore, surgical intervention, if needed, has to be carefully planned, and these softwares can help in virtual simulations of interventions, as well as in actual conductions of operations [8].

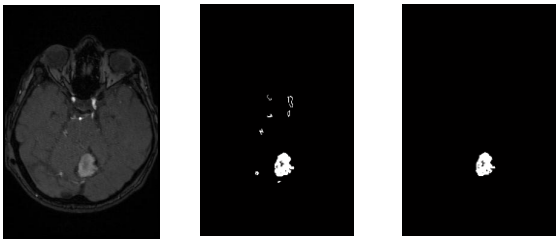
In the Computer Assisted Surgery system, the actual intervention is defined as a surgical navigation. It consists of coordinated actions of a surgeon and the surgical robot, which is pre-programmed to perform tasks during the operation. The surgical robot is a mechanical device, which usually has the appearance of a robotic arm and is computer driven. Using neuronavigation system as a tool, the surgeon can see the instruments on the computer and to confirm their exact location at any time. Analyzing the point to which the instrument is pointing in the patient's body, and then the monitor, the surgeon makes a connection between the body and the screen image. By comparing the scanned image with the real anatomy, the surgeon during the operation, using the instruments guided image, can see the location of the tip of the instrument in the patient's body and its performance [9].

III. MATERIAL AND METHODS

In this study 497 CT images of a brain from a real patient were used. Images were divided in four different categories according to the view - axial, coronal, reformatted coronal, and reformatted sagittal images. Images were represented by different matrixes, out of which axial images were represented by 384 x 256 pixel matrix. There were 133 axial images on the whole which were used for application of algorithms presented in this paper.

This paper presents algorithms to differentiate the brain tumor from healthy tissue based on the pixel difference. For easier operation with the matrix, each image was transformed into a matrix with a range of intensity or gray level value of a pixel from 0 to 1, where 0 corresponds to the completely black background, and 1 completely white.

Furthermore, a certain range of gray level value of the tumor in the image is declared as well as minimum size (area) of the tumor is defined, in order to avoid recognition of the parts of the brain that are within a defined level range of gray intensity, but are not a tumor (Figure 1).



a) CT image b) isolated tumor without size limitation c) isolated tumor with size limitation

Figure 1. Comparison of CT scan codes with efficiency in tumor isolation

Separation of the tumor was performed using a recursive procedure depth first search - DFS. Generally speaking, DFS algorithm is used for traversing or searching tree or graph data structures [10]. One starts at the root (selecting some arbitrary node as the root in the case of a graph) and explores as far as possible along each branch before backtracking. Algorithm starts at the root of the tree (graph at some node to determine the root), and then crawls along all the branches as much as possible before returning to the root. The tour starts from a given arbitrary node r, going further in depth. The root is marked as visited. Then an arbitrary unmarked node r_1 adjacent to r is selected and the recursive procedure searches in depth. When a procedure comes to a node whose all neighbors (if there are any) are already marked as visited, the recursion procedure exits that level. If at the point of the completion of a search of r_1, all neighbors of node r are labeled as visited, then the search for a node r ends. Otherwise, the next unmarked neighbor r_2, arbitral to a node r is selected, and the search starts from r_2, etc. This procedure is combined with the storage of pixels matrices which are visited and within the defined level of gray intensity. Those pixels that meet the given conditions are colored in white (they are assigned a logic 1), whereas those which do not meet the conditions are colored in black (they are assigned a logic 0). These members are also referred to as visited from that point.

Further analysis considers the size of the surface colored in white. In accordance to the minimum defined size of a tumor, a set of white connected pixels with the region area greater than minimum are left white (they represent tumor tissue), and those that are smaller than the defined size are colored in black using the same method as in previous DFS procedure.

Algorithm for tumor recognition and its extraction

Step 1: Convert DICOM image into matrix with numbers in range from 0 to 1

Step 2: For each pixel determine if it is in defined range of gray intensity:

If pixel is in defined range, set its value to 1

Else, set its value to 0.

Move to next pixel

Step 3: Mark all pixels as unvisited

Step 4: For each pixel do:

Check if the pixel is white and not visited

Start DFS procedure for that pixel

If surface of neighboring white pixels is less,

than defined surface area, color it in black

Step 5: Show extracted image with white tumor on a black surface

Segmentation is implemented for 133 axial view CT slices. These segmented slices were stacked as new images for tumor volume rendering.

It was of great interest on the original image of a brain with a tumor to extract a tumor and distinct it visibly from other tissue (in this case color it in red). Using previously mentioned algorithm and combining it with the fact that RGB (red-green-blue) image representation means that grey image (three-dimensional matrix) is represented with same amount of red, green and blue presence, it was possible to color extracted tumor in one color (red). By setting two other colors (green and blue to 0) and red to 1 for the pixels that meet the specified conditions and doing nothing if the conditions are not met, the tumor was visibly distinguished from surrounding tissue.

Algorithm for coloring tumor on an original image

Step 1: Create new matrix

Step 2: For each pixel check:

If a pixel from an image with extracted tumor is black do:

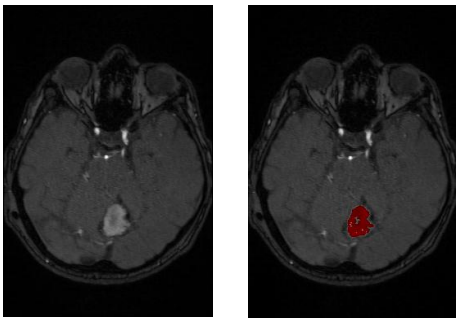
Leave red, green and blue component the same as in the original DICOM image

Else, leave just red component (assign 1 to red component and 0 to green and blue)

Step 3: Convert matrix to the one with numbers from 0 to 1

Step 4: Show image with colored tumor in red

The result of the previous algorithm is given in the figure 2.



a) original CT image b) CT image with extracted tumor

Figure 2. Original CT image of a brain in comparison to the CT image with extracted tumor colored in red

Limitation of the algorithm presented in this paper is that depending on the complexity of tumor, critical surface area has to be determined in order to achieve best possible results. In cases of complex shapes of cross-sections of tumor where two or more areas of cross-section of the same tumor are present, surface limitation has to be set for smaller values in order to include both smaller and greater surfaces of cross-section of tumor. It is possible in those cases, because of small surface area limitation that some parts are recognized by computer a as tumor, although they are not a tumor. This leaves a possibility for improvements in presented algorithms.

IV. RESULTS

A 3D model of the brain based on CT images was made, using only images that represent the axial view. Firstly, it was necessary to choose from a series of images those that represent axial view (separated from coronal, reformatted coronal, and reformatted sagittal view) and then execute the code considering putting images as layers of a model in the exact order where each image has a precise depth defined by data that is stored in the image (Figure 3). Axes and model name are set; all images are defined as gray.

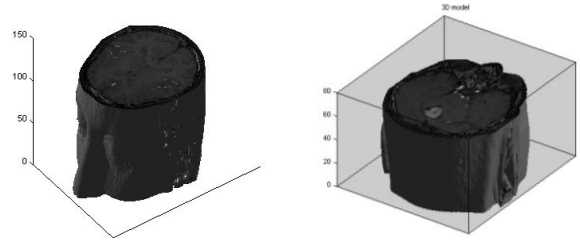


Figure 3. A three-dimensional model of the head based on CT images of a real patient

After segmentation of a tumor and making 3D model of a brain, the same method that considers putting images as layers of a model in the exact order where each image has a precise depth defined by data that is stored in the image is used to make three-dimensional model of a tumor. As a part of pre processing, tumor extraction was necessary in order to avoid the misclassifications of the surrounding tissues. By removing non-tumor tissues, only tumor tissues will remain in the image. These images are stored and used for rendering 3D model of a tumor. Some of the views on the tumor segmented from DICOM images are shown on Figure 4. As it can be seen from the scale on the Fig.4, tumor is located between first and eightieth CT slice.

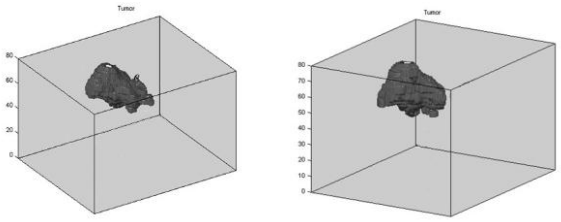


Figure 4. A three-dimensional model of segmented brain tumor

V. DISCUSSION

The extraction of 3D objects and their visualization is one of the most important steps in the analysis of the preprocessed medical image data, which can help in performing diagnosis, treatment planning, and treatment delivery. Thus in practice, radiation oncologists spend a substantial portion of their time performing the segmentation task manually, using one of the available visualization and segmentation tools [5].

The goal of this automatic segmentation and 3D modeling tool was to make automatic segmentation and 3D modeling of CT images more practical than using manual segmentation that is common at radiation oncologists. It is very difficult to measure the performance of enhancement objectively. If the enhanced image can make observer perceive the region of interest better, then we can say that the original image has been improved.

Volumetric tumor representations are suitable for image registration, surgical planning for detection of tumor growth and also for neuronavigation in cases of operation. Conventional systems of neuronavigation still have practical limitations owing to the lack of an intraoperative imaging modality to provide the surgeon with information regarding dynamic changes that occur during surgery. The first step in enabling neuronavigation become present in medical operations is to enable automatic recognition and segmentation of a brain tumor. After the models are made and integrated with operational instruments management, it is possible to make navigation system with high-quality, real-time, three-dimensional (3D) imaging capabilities [11]. Intraoperative CT, in which the surgeon operates the patient and has the overview in real time of what is happening inside the brain through two-dimensional (2D) images as well as 3D model with which he can manipulate without moving the patient [12].

VI. POSSIBLITY OF MAPPING SUCH ALGORITHMS ON HIGH PERFORMANCE RECONFIGURABLE COMPUTERS

Depending on the image quality and number of slices, the process of segmentation and 3D visualization of a brain with tumor can be very slow. Volumetric rendering of tumor volume using manual segmentation of its outlines can be a time-consuming process. In those cases the utilization of High Performance Reconfigurable Computers (HPRCs) can be very useful.

The architecture that integrates the reconfigurable computers (RC) and general-purpose processors or some parallel computing systems is called High-Performance Reconfigurable Computers. The RC systems are in most cases based on Field Programmable Gate Array (FPGA) architecture and they represent the combination of reconfigurable logic. Field programmable gate arrays represent integrated circuits which are designed in such manner so that they can be configured by the designer in the purpose to implement some digital computations. An example of RC system is FPGA based Dataflow Engine (DFE). The DFE contains FPGA as a computational engine and can be easily integrated with some host system which can be a general-purpose processor or some parallel computing system. The usage of HPRCs can achieve great increase in calculation speed in many algorithms.

One example of DFE is Maxeler's MAX2336B accelerating card. It is attached to the host processor via PCI Express bus and it is configured by the designer with one or more kernels and a manager. Kernels are the parts which represent the hardware implementation of some algorithm. The manager has as a main task to define dataflow between kernels, on chip memory and host processor.

By integrating the Maxeler's DFE with some host computer we get the Maxeler dataflow computer. It can be understood as combination of two programming paradigms: control flow and dataflow. Before one begins programming DFE, he must split the whole algorithm into two parts: control flow and dataflow. The control flow part is the part which will be executed on host computer, while the dataflow part will be executed on DFE.

The algorithms such as the algorithm for tumor recognition and its extraction and the algorithm for coloring tumor on an original image which are described in this paper can be easily mapped on Maxeler dataflow computer. The control-flow part would be the part which loads the images and sends them pixel by pixel to the DFE for segmentation, and after the segmentation collects pixels from the DFE and stores them on hard drive in appropriate format. The dataflow part would be the part which executes the main calculation, recognizes the tumor and extracts it or colors it. Large speedups can be achieved by mapping these algorithms on Maxeler dataflow computer.

VII. CONCLUSION

Relevance of techniques presented in this paper is the direct clinical application for segmentation. In this study from a set of 497 CT images in standard DICOM format, 133 CT images which represented axial view were selected for further processing. We wrote program that localizes brain tumor and extracts it on every DICOM slice. We have described method of segmentation and extraction of a brain tumor in medical image processing and discussed properties of techniques in brain tumor detection. This paper is used to give more information about brain tumor detection and segmentation. The target area is segmented tumor which was used to make a three-dimensional model as well as three-dimensional model of a brain.

In future, the system should be improved by adapting more segmentation algorithms to suit the different medical image segmentation cases and more precise localization. Combination of 3D models can be helpful to doctors in diagnosis, the treatment plan making and state of the tumor monitoring. Further development of the program would involve the determination of the optimal trajectory approach to the tumor with the least possibility of damage to functionally important parts of the brain. The application of those programs would lead to their application in neuronavigation and neurosurgery. The use of such programs would involve creation of 3D models and manipulation with the model, as well as allowing an insight into the position of the instrument at any time. Thus, a surgical procedure in which the surgeon uses a surgical instrument tracking with preoperative or postoperative images that would indirectly lead procedure, would greatly facilitate the execution of the sensitive operation to remove a brain tumor. Programs that would enable performance of such complex operations would integrate medical knowledge with the engineering and require working with programming languages of high precision.

ACKNOWLEDGMENT

The part of this research is supported by Ministry of Science in Serbia, Grant III41007 and ON174028.

REFERENCES

[1] NR Pal, SK Pal. "A review on image segmentation techniques," *Pattern Recognition* vol 26, pp.1277-1294. 1993

[2] T. Logeswari and M. Karnan, "An improved implementation of brain tumor detection using segmentation based on soft computing", *Journal of Cancer Research and Experimental Oncology* Vol. 2(1) pp. 006-014, March, 2010

[3] National Brain tumor society , www.braintumor.org

[4] Ho, S., Bullitt, E. and Gerig, G. (2002) Level-set evolu- tion with region competition: Automatic 3-D segmenta- tion of brain tumors. *Proceedings of International Con- ference on Patter Recognition*, Quebec, 11-15: 532-535., August 2002.

[5] Robb, R., Hanson, D., Karwoski, R., Larson, A., Work- man, E. and Stacy, M. Analyze: A comprehensive, operator interactive software package for multidimensional medical image display and analysis. *Computerized Medical Imaging and Graphics*, 13: 433-454., 1989.

[6] J Chunyan, Z Xinhua, H Wanjun, M Christoph, "Segmentation and Quantification of Brain Tumor", *IEEE International conference on Virtual Environment, Human-Computer interfaces and Measurement Systems*, USA pp. 12-14., 2000.

[7] KS Fu, JK Mui "A survey on image segmentation", *Pattern recognition* 13: 3-16. 1981.

[8] G. Unsgaard, S. Ommedal,T. Muller, A. Gronningsaeter, T.A.N. Hernes, "Neuronavigation by Intraoperative Three-dimensional Ultrasound: Initial Experience during Brain Tumor Resection", *Neurosurgery*, Vol. 50, No. 4, April 2002

[9] CR Wirtz, FK Albert, M Schwaderer et all. "The benefit of neuronavigation for neurosurgery analyzed by its impact on glioblastoma surgery". *Neurol Res* 22:354–360, 2000.

[10] Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to algorithms*. Cambridge: MIT press, Vol. 2. Pp 549., 2001.

[11] H Gumprecht, CW Darius, CB Lumenta, "Neuronavigation System: Technology and clinical experiences in 131 cases". *Neurosurgery* 44:97–105, 1999.

[12] PMcL Black, T Moriarty, E III Alexander, P Stieg et all. "Development and implementation of intraoperative magnetic resonance imaging and its neurosurgical applications" *Neurosurgery* 41:831–845, 1997.

Framework for early manufacturability and technological process analysis for implants manufacturing

Miloš Ristić*, Miodrag Manić**, Boban Cvetanović*
* College of Applied Technical Sciences Niš, Niš, Serbia
** University of Niš / Faculty of Mechanical Engineering, Niš, Serbia
milos.ristic@vtsnis.edu.rs, miodrag.manic@masfak.ni.ac.rs, boban.cvetanovic@vtsnis.edu.rs,

***Abstract*— Patient specific implants, i.e. custom designed implants for an individual patient, are growing in popularity. This paper presents concept of a system, as well as the methodology of manufacturability analysis for custom designed implants manufacturing. System uses 3D model of implants with integrated implant knowledge. It also integrates a knowledge base and processes data base used for implants manufacturing. Set of rules gives recommendations for implants manufacturing and ranks them.**

I. INTRODUCTION

Contemporary approaches in integral development of a product implies methodology of simultaneous/concurrent engineering based on joint work of specialist experts in certain areas in the earliest phases of product inventing and designing. In the early phases of virtual product development, various experts' knowledge should be implemented in order to reduce development time and product price. Thus expert knowledge and experience is implemented in order to be able to completely observe, visualize and test it for work in simulated exploitation conditions [1, 2].

Through virtual product development various aspects of that product are considered, starting from conceptual design to its behavior in practice. One of the important phases in product creation is deciding on technologies and procedures for the most optimal product realization. It means that various methods and technologies should be considered and the most optimal one chosen. Thus, time of product realization and product price would be significantly reduced.

This paper presents concept of a system, as well as a methodology for manufacturability analysis for manufacturing implants used in medicine in surgical interventions. The paper is focused implants that are not typical and standard but are adjusted to patient specific needs (customized implants) [3]. Patient specific implants, i.e. custom designed implants for an individual patient, are growing in popularity.

The geometry and topology of those implants are adjusted to the anatomy and morphology of the selected bone of the specific patient. Their application has a positive effect on patients, but on the other hand requires more time for preoperative planning and manufacturing. Therefore, these are used in areas where the application of predefined fixators can lead to complications in the surgical interventions or in the process of recovery [4].

The second chapter of the paper describes reversible engineering customized implant design process [5]. Further on, the paper describes Knowledge Based (KB) implant manufacturability analysis, which gives a basis for manufacturability process analysis for manufacturing dynamic fixation on Tibia [6]. Decision on the optimal procedure choice is made by ranking criteria of the most important characteristics of a procedure (time, accuracy and cost).

II. DESIGN OF CUSTOMIZED IMPLANTS

An implant is a medical device manufactured to replace a missing biological structure, support a damaged biological structure, or fix an existing biological structure.

Conventional implant manufacturing methods provided that certain parts are manufactured in standard defined range. This way, certain implants couldn't adequately respond to patient specific needs, and the post-operative recovery was more difficult.

The creations of 3D models of customized implants are based on the 3D models of the patient specific bones. For this purpose, the first step is to create a 3D model of a selected bone [7]. Typical design loop is shown in Figure 1.

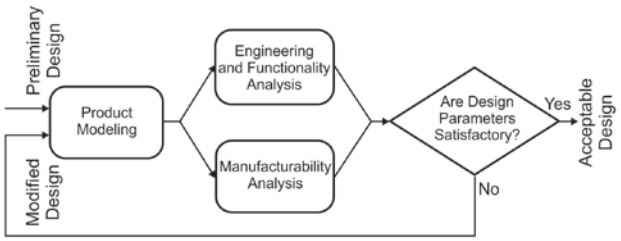


Figure 1. Typical design loop.

The creation of geometrically accurate 3D models of human bones utilizes number of different techniques and presents a unique challenge, because their geometry and form are very complex. These types of shapes can be modeled by using surface patches represented by Bezier or B-spline surfaces, or by using NURBS patches, which are commonly used in traditional CAD applications, e.g. CATIA [8].

The output of CATIA is presented in STL (Stereolithography) format, which allows it to be directly transferred into an RP (Rapid Prototyping) machine.

Reverse engineering of human bones implies the use of some kind of medical imaging device for the acquisition of medical data (Computed Tomography – CT, Magnetic

Resonance Imaging – MRI), then processing that data in medical or CAD software, and at the end, creating a valid geometrical model (surface, volume) [5].

The general reverse engineering techniques for design of customized implants, for selected bones, include several tasks (Figure 2).

- Importing and editing point cloud acquired from medical imaging device,
- Tessellation of point cloud and creation of polygonal model (mesh),
- Anatomical and morphological analyses of a selected bone,
- Identification of RGEs (Referential Geometrical Entities) which are based on the anatomical and morphological characteristics of a selected bone (points, directions, planes and views) [9],
- Creating and editing the curves on polygonal model of the selected bone, in accordance with the RGEs,
- Creating and editing the surface model of the selected bone's outer surface by sweeping, lifting, blending, and trimming the curves,
- Selection the places on the selected bone where the implant will be placed,
- Adjusting the geometry of the implant according to the requirements of the surgeon,
- Creation and modification of 3D models of implants,
- Simulation of implant placement to selected bones.

This methodology gives the opportunity to create a custom implant design adjusted to the patient's anatomy, improving structural, functional and aesthetic biocompatibility.

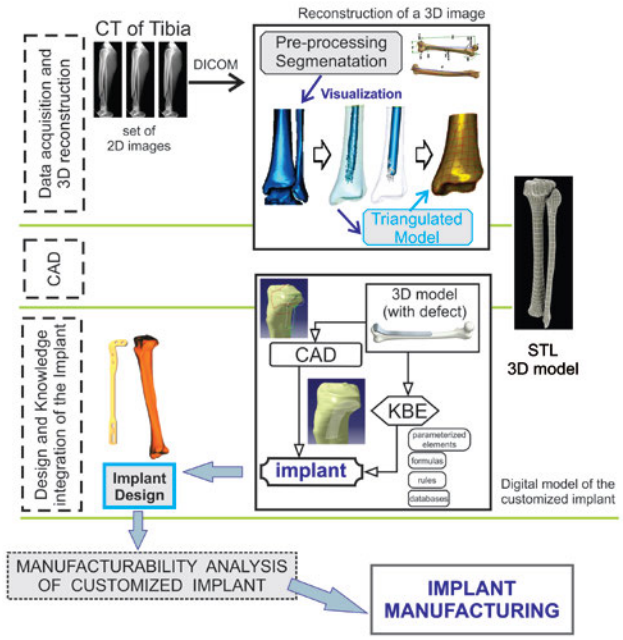


Figure 2. Automated design process phases and manufacture of customized implants

Implant modeling is performed in Computer-Aided development systems (CAPD), which, beside geometry and topology, enable integration of product knowledge, and manufacturing possibilities and applied technologies restrictions for some forms, in a virtual model of a product.

This knowledge is used for manufacturability analysis of suitable implant manufacture processes. In order to decide on the customized design for final user, some geometrical rules and restrictions can be included in 3D model, by means of databases, like "thickness" or „available tools for implant manufacturing".

Further on, the paper describes some of the existing methods and possible technological procedures and processes for manufacturing customized implants with their most important characteristics, advantages and disadvantages. Thus, a framework for manufacturability process analysis of specific implant is set.

III. MANUFACTURABILITY ANALYSIS OF COSTUMIZED IMPALNT

Nowadays, in medicine a great number of different biocompatible materials for implants manufacturing are used. There are also a lot of manufacturing and material processing technologies, starting with conventional technologies, CNC technology to additive technologies. All of them have their own characteristics and influence the quality and price of the product. It is very important to choose an adequate technological procedure which gives implants satisfying exploitation characteristics, has optimal cost and short manufacturing lead time. This is very important for customized implants.

Even during the implant design phase it is important to consider and analyze possible technological procedures. That analysis is provided by using manufacturability analysis and analysis of systems and applications.

A product is manufacturable if it is suitable for manufacturing with planned technology. Therefore, when designing a product we try to find a solution that requires minimum use of manufacturing time and material with minimum necessary equipment for manufacturing a product suitable for a particular purpose and function. Product manufacturability as a production convenience measure is a very broad term and it is difficult to unilaterally define, because it depends on many influential factors, including process conditions. In the manufacturability analysis there is no absolute measure. What is manufacturable for one product may not be for the other [10].

According to the approach to considering manufacturability these systems could be divided into direct approach systems based on rules and checks; and indirect approach systems based on generating manufacturing plan and procedure, and then on modification of various procedures in order to reduce costs.

Measures of manufacturability for assessing the level of manufacturability:

- Binary measures (0 or 1/ yes or no/ manufacturable or non manufacturable/ …);
- Qualitative measures for descriptive manufacturability measures of virtual prototype (poor, average, good, excellent);
- Abstract-quantitative gives a manufacturability level by assigning numerical values to the abstract scale (e.g. assigning a manufacturability index range

between 0 and 1 and their combining in the final grade by using methods such as Fuzzy logic);

- Time and cost as a manufacturability measure are two most important parameters of a technological process easily combined in summative manufacturability rating, but they can not directly help designer in the assessment whether he really achieved satisfactory level of product manufacturability.

Depending on the moment of manufacturability analysis, we can define two approaches: analysis during the design process (on-line); and manufacturability analysis upon the design process completion (off-line).

Manufacturability analysis systems are actually Knowledge based systems (KB). Certain CAD programs integrate KB systems, thus CATIA has a "Knowledgeware" module, which is a kind of an expert system. In Knowledgeware, analysis is enabled by integrating information and data which are through certain relations (specified in VBScript) connected in model knowledge (e.g. parameterize elements, databases, create formulas or rules, check, etc.).

IV. IMPLANT MANUFACTURABILITY PROCESS ANALYSIS

Manufacturability process analysis actually assesses technology applicability by comparing techniques for creating certain designed implant and defines the techniques achieving maximum level of required quality with minimum costs and manufacturing lead time.

Generally, all the techniques, according to manufacturing processes, can be divided into three basic groups:

- *Formative Manufacturing are processes of material shaping by solidification- consolidation process;*
- *Substractive Manufacturing (Machining) is process where piece of raw material is cut into a desired final shape and size by a controlled material - removal process; and*
- *Additive Manufacturing are processes which are based on connecting points or material layers with the aim of achieving a desired shape- 3D printing process.*

In order to recognize a particular technological procedure and analyze a level of its applicability, Table 1 gives a review of certain procedures and some of their characteristics used to decide on the optimal process. Considering the fact that there is a great number of procedures, techniques and methods for manufacturing a particular implant, the table presents only processes related to customized implant manufacturing procedure on the example of dynamic tibia fixation.

Manufacturability process analysis is seen in establishing the level of its applicability (if the process is applicable). And on the other hand which characteristics make is inapplicable (if the process is inapplicable).

Measures for establishing the level of technique applicability defined in Table I are: Accuracy, Cost and Manufacturing Lead Time. A value (weight) was assigned to each of these measures showing the importance of that value for establishing measures for manufacturability process analysis.

TABLE I.

Review of established manufacturing techniques used in medical field

	Accuracy	Cost	Manufacturing Lead Time
Stereolithography (SLA)	+++	$$	+++
Selective Laser Sintering (SLS)	++	$$$	+++
Fused Deposition Modeling (FDM)	++	$	+++
Direct Metal Laser Sintering (DMLS)	++	$$	+++
3DP	+	$	+++
Milling	++	$$$	+
Turning	+	$$$	+
Injection Molding	++	$$$$$	+
Sand Casting	++	$$$$$	+
Forging	+++	$$$$$	+

Conceptual scheme of a system for manufacturability process analysis is shown in the figure 3. System Input data on one hand are received from an implant model data base. System on the other hand involves Process Capability Database, in which available processes are presented with a set of rules and application restrictions related to constructive - technical features integrated in an implant model.

By applying rules system analyzes 3D implant model and gives recommendations for its manufacturing. It also ranks recommended procedures in accordance with knowledge stored in process knowledge base.

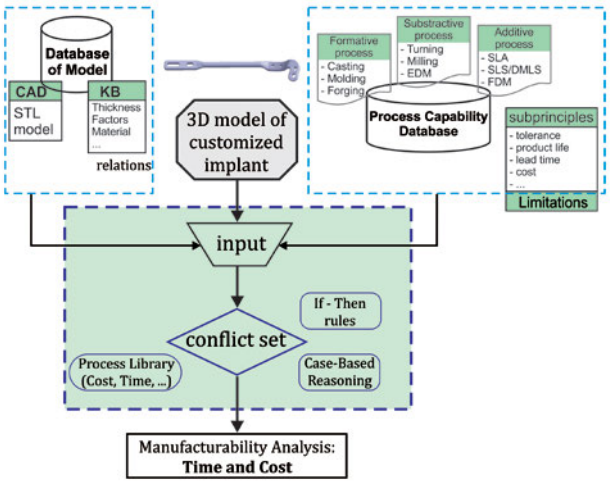


Figure 3. Concept of Manufacturability Process Analysis

The system is based on the use of If-Then rules. Beside If-Then rules, the system can use other available recommendations such as for example Case-Based decision making or similar known process libraries. Thus technological procedure knowledge about manufactured implant can be used for later comparisons of process criteria. This becomes particularly useful if a certain manufactured implant, primarily in its geometry and its position (on the bone itself), but also in other characteristics, is similar to customized implant model being manufactured. Existence of such data enables earliest design and process manufacturability analysis.

When analyzing, following criteria can be set: cost, manufacturing time and accuracy, or some other. These three criteria are used in manufacturability analysis of internal dynamic tibia fixation according to Mitkovic type TPL, (Figure 4) [5, 8, 11].

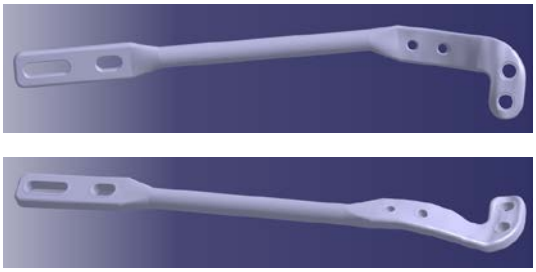


Figure 4. 3D Model of dynamic fixation according to Mitkovic

Analysis results can be reduced to the following conclusions, i.e. recommendation.

The shortest fixator manufacturing lead time and the most accurate shape are provided using additive manufacturing. But, the cost is by far the highest.

The cheapest fixator can be manufactured using classical manufacturing technologies (cutting and deformation), but this process is the longest and does not always give the same quality.

The most optimal procedure is the forging process which gives both good quality and reasonable price, but requires utilization of significant tool manufacturing resources, and it is the most optimal method for manufacturing large series of a product, which is rare in customized implant production.

V. CONCLUSION

This paper presents the concept of the system and the initial result for manufacturability analysis for customized implants. The system uses a parameterized 3D model of customized implants, which integrates general and expert knowledge about particular elements of implant construction. The knowledge is gathered from material manufacturers, production equipment manufacturers, orthopedists, implant manufacturers, and other experts involved in its realization. System is based on the use of knowledge base and rule base for implants, materials and manufacturing techniques. It is implemented in CATIA through Knowledgeware module.

Its application it the implant design phase gives recommendations and ranks potential technological procedures.

System is designed as an open system for upgrading knowledge base and rule base and gives a good basis for development of quality and applicable system for practical use.

ACKNOWLEDGMENT

This paper is a result of the project III 41017, supported by the Ministry of Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] Y.-S. Ma, · G. Chen, and · G. Thimm, “Paradigm shift: unified and associative feature-based concurrent and collaborative engineering”, *Journal of Intelligent Manufacturing*, Vol. 19 Issue 6, pp 625-641, December 2008.

[2] M. Manic, V. Miltenovic, M. Stojkovic. And M. Banic, “Feature Models in Virtual Product Development”, *Strojišni vestnik – Journal of Mechanical Engineering*, vol. 56 (3). 2010. pp. 169-178.

[3] V. Chulvi. D. Cabrian-Tarrason, A. Sancho, and R. Vidal, “Automated design of customized implants”, *Rev. Fac. Ing. Univ. Antioquia* N.° 68 pp. 95-103. Septiembre, 2013

[4] J. Arnone, *A comprehensive simulation-based methodology for the design and optimization of orthopaedic internal fixation implants*, Ph. D. Thesis, The Faculty of the Graduate School, University of Missouri-Columbia, 2011.

[5] V. Majstorovic, M. Trajanovic, N. Vitkovic, M. Stojkovic, “Reverse engineering of human bones by using method of anatomical features”, *CIRP Annals - Manufacturing Technology,* Vol. 62, pp. 167–170, 2013.

[6] D. Stevanovic, N. Vitkovic, M. Veselinovic, M. Trajanovic, M. Manic and M. Mitkovic, Parametrization of internal fixator by Mitkovic, *International Working Conference Tottal Quality Management – Advanced and Intelligent Approaches*, pp 541-544, Belgrade, Serbia, June 2013.

[7] Thomas, Thaddeus Paul. "*Virtual pre-operative reconstruction planning for comminuted articular fractures*." PhD thesis, University of Iowa, 2010. http://ir.uiowa.edu/etd/2778

[8] N. Vitković, M. Veselinović, D. Mišić, M. Manić, M. Trajanović and M. Mitković, “Geometrical models of human bones and implants, and their usage in application for preoperative planning in orthopedics”, *11th International Scientific Conference MMA 2012 – Advanced Production Technologies*, pp. 539-542, Novi Sad, 2012.

[9] N. Vitković, M. Trajanović, J. Milovanović, N. Korunović, S. Arsić, and D. Ilić, “The geometrical models of the human femur and its usage in application for preoperative planning in orthopedics”, *ICIST 2011*, Kopaonik, Serbia, 2011.

[10] M. Ristic, “*Product designing in terms of manufacturability*”, Master thesis, Mechanical Engineering Faculty of the University of Niš, Niš, 2012.

[11] M. Mitkovic, S. Milenkovic. I. Micic, D. Mladenovic and M. Mirkovic, “Results of the femur fractures treated with the new selfdynamisable internal fixator (SIF)”, *European Journal of Trauma and Emergency Surgery*. Vol. 38 (2), pp. 191-200, April 2012.

Multimodal Imaging for PET Attenuation Correction

Nikola Mijailović[1], Jasna Radulović [1], Miroslav Trajanović [2], Nenad Filipović [1], Aleksandar Peulić[1]
[1] Faculty of Engineering, University of Kragujevac, Serbia
[2] Faculty of Mechanical Engineering, University of Niš, Serbia
nmijailovic@kg.ac.rs, jasna@kg.ac.rs, miroslav.trajanovic@masfak.ni.ac.rs, fica@kg.ac.rs, aleksandar.peulic@kg.ac.rs

***Abstract*—In this study different methods of medical imaging are considered along with the possibility of combining information offered by them. One of the techniques of special interest is positron emission tomography (PET). This medical imaging technique gives information about metabolic processes in the human body. One of the main defiance of this method is attenuation of gamma photon in interaction with tissues and organs inside the body. This phenomenon can be fixed by the attenuation correction factor. The factor can be obtained using Computerized tomography (CT) imaging with aim to map PET image with linear attenuation coefficient. The PET and CT images are matched using image registration technique which provides determination of attenuation coefficient space distribution. Using the attenuation coefficient the pixels intensity of PET image can be scaled by the reciprocal value of attenuation coefficient with propose of attenuation correction.**

I. INTRODUCTION

The medical imaging is a basic technique in diagnostics and decision making in medicine. The medical imaging methods based on the obtained information about human body primarily use electromagnetic radiation. The large progress in medical diagnostic has been made with the development of nuclear medicine. The nuclear medicine is based on using radio nuclide which emits product of nuclear decay, usually gamma quant, which can be detected by the detectors. PET is an important imaging method in nuclear medicine area [1]. This technique uses positron annihilation process for detecting some metabolic processes in human body. This method, when the radiation source is positioned into the patient body, is one of the emission techniques. A radiation technique uses radiation source outside the human body. Emitted radiation interacts with tissues and organs and its attenuation can be obtained by detector. CT imaging is the most common radiation technique where the radiation source is X-ray tube [2]. The Magnetic Resonance Imaging (MRI) is technique complementary to CT which gives information of tissue structures using external magnetic field and radio frequency radiation [3]. Using only one medical imaging technique in many cases can not give sufficient data for medical decision support. This fact causes the need for using multimodal imaging method.

II. MULTIMODAL MEDICINE IMAGING

The multimodal imaging means the combination of the different methods of medical imaging with the aim to obtain more quality information of the human body.

A. Computerized tomography (CT)

Computer tomography is a standard radiology method for 3D object scanning using X-ray radiation. This method is based on the differences in attenuation coefficient of X-ray beams for various materials and tissues. The final result is a grey level CT image where corresponding grey level is proportional to attenuation coefficient [4].

CT medical imaging includes exposure of the object of radiation at one side and detecting attenuated radiation at the other side of the object and this procedure is repeated from more than one direction. The next step is image reconstruction from the projection by using a number of techniques. All of these techniques are based on solving systems of integral equations which are formed as a result of total attenuation of the radiation beam from the source to the detector. Attenuation of X-ray radiation in the homogenous media with linear attenuation μ is given by relation:

$$I = I_0 e^{-\mu \cdot d} \tag{1}$$

where I_0 is intensity of initial radiation, I is final intensity radiation after path length d in tissue with linear attenuation coefficient μ. If there are multiple materials, the equation becomes:

$$I = I_0 e^{-\sum_i \mu_i d_i} . \tag{2}$$

Linear attenuation coefficient depends on X-photon energy; due to equations 1 and 2 are satisfied only for monochromatic beam.

If a polychromatic X-ray source is used, taking into account the fact that the attenuation coefficient is a strong function of X-ray energy, the complete solution would require solving the equation over the range of the X-ray energy (E) spectrum utilized:

$$I = \int I_0(E) e^{-\mu(E) \cdot d} dE \tag{3}$$

In CT practice examination the CT numbers also known as the Hounsfield units are used. The CT numbers are proportional to mean linear attenuation coefficient [5]. The CT number is defined as

$$CT = K\frac{\mu_t - \mu_w}{\mu_w} \quad (4)$$

where μ_t, μ_w are linear attenuation coefficients of X –ray in tissue and water respectively and K is scale factor. The K is chosen to satisfy -1000 value of CT number for air and 1000 for some kind of cortical bone tissue. The CT imaging provides high resolution and satisfactory separation between soft tissue and bone. The disadvantages of this medical imaging method are exposure dose of the X-ray radiation what is potentially risk for health safety and low contrast in soft tissue examination.

B. *Nuclear magnetic resonance imaging (MRI)*

Nuclear Magnetic resonance is a spectroscopic method of imaging in medicine. This method is based on the measurement of the proton concentration (hydrogen atom) in tissue structures and organs [6].

The hydrogen protons of human body tissue rotate around its own axis. These protons take free orientation in the space and total magnetization is closely equal to zero. In presence of the external magnetic filed the atom nucleus is oriented in the magnetic field direction with Larmor frequency:

$$\omega = \gamma B \quad (5)$$

where B is applied magnetic filed and γ=21432324 is gyro magnetic ration. The magnetic field value is between 0.5 T to 2.5 T.

When radio-frequency pulse is applied on the human body, the magnetic moments of the nucleus are tilted by certain angle toward the initial position.

At the moment when radio frequency is removed the nucleuses lose energy by the emitting certain radio frequency, this signal is referred to as the free-induction decay and this is referred as relaxation.

Free-induction decay response signal is measured by a conductive field coil placed around the object being imaged. According to measurement data the 3D grey-scale MR images can be reconstructed. This signal value represents the nuclei concentration density. For 3D reconstruction the radio frequency signal needs to be encoded for all directions. Axel direction is obtained using gradient magnetic field which causes change of Larmor frequency in that direction.

MR image contrast also depends on the two other tissue-specific parameters: measures the time required for the magnetic moment of the displaced nuclei to return to equilibrium. The time T1 refers to the period necessary for returning the magnetic moment in the direction of applied magnetic filed, and T2 is time required for the magnetization vector normal to applied field becomes zero.

C. *Positron emision tomography (PET)*

The positron emission tomography is an emission tomography based method of imaging. This method is based on the creation of positron in nuclear decay and their annihilation with free electron [1]. The typical radionuclide used for creation of positron is fluorine 18 or carbon 11. As the result of this interaction the pair of gamma photon is created every per 511 KeV of energy.

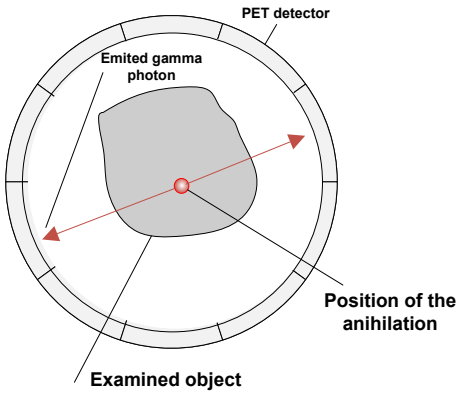


Figure 1. Principle of the PET

These results are a consequence of the energy conversion total mass of electron and positron into gamma photon energy. The photons have opposite direction due to conservation of the momentum. The emitted gamma photon can be detected on the detector positioned oppositely at different side of the observed object, Figure 1. The 3D image of an object can be obtained by reconstruction procedure of the measured dose in the detector. The detectors are placed at the end of the photon way and this scheme is called the coincidence detection. The position of annihilation in the human body is measured according to time delay between two detections in the opposite side detector. This time is on the order of nanoseconds. The condition for detection is presence of both events of detection inside defined time window. The common type of detector is bismuth germanate oxide.

III. ATTENUATION CORRECTION IN PET STUDY

Photon attenuation is the process of interaction with atoms and other particles resulting in a complete photon absorption or scattering with energy loss [7]. The percentage of photons attenuated within the tissue is independent of the annihilation location, but it is dependent on the total travel length of the two 511 keV photons along a line-of-response (LOR) [8].

In the Figure 2 the attenuation of emitted gamma quant is shown. The number of photon can be absorbed or scattered in the tissue and as a results the information about tissue properties is lost. The attenuation of particle beam is dependent on the attenuation of tissue and line of response value.

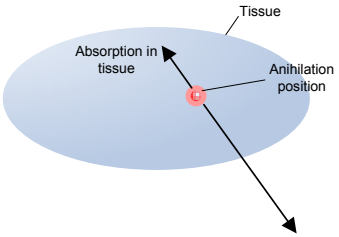


Figure 2. Attenuation of gamma photon in PET study

This dependency is given by relation.

$$A = A_0 e^{-\mu \bullet lor} \tag{6}$$

where A_0 is total number of annihilation at given point in certain direction, μ is attenuation coefficient along path of photon trajectory with corresponding *lor* value and *A* is total number of registered photon in the detector [9].

Basic idea of attenuation correction is measure or calculation of μ(x,y) and *lor* for given point with coordinate x and y and multiply the number of registered photon by the factor of attenuation correction which is given by

$$AC = e^{lor \bullet \mu\_tissue} \cdot e^{lair \bullet \mu\_air} \tag{7}$$

where the $e^{lair \bullet \mu\_air}$ is air attenuation influence. [10]

A. CT based attenuation correction method

The attenuation correction based on the CT is method to directly measure the linear attenuation coefficient of the tissue and organs. In a more precise consideration CT images represent map of effective linear attenuation coefficient for the energy range interval given by the spectra of the X-ray source. For the attenuation correction we need to convert this value to the linear attenuation coefficient for the energy of the 511 KeV. [11] The linear attenuation coefficient strongly depends on the energy and there is not an unambiguous method which fully resolves this problem in emission tomography imaging. The gray value of CT image is corresponding to CT number defined in relation (4). The CT number is directly proportional to the linear attenuation coefficient of the tissue. Using the linear interpolation attenuation coefficient for given CT X-ray spectra can be convert to linear attenuation for photon energy of 511 KeV. Based on this calculation the space distribution of the linear attenuation coefficient can be obtained. Knowing the distance of the given pixel to the detector provides an opportunity for attenuation correction using relation (7).

B. MRI based attenuation correction method

The MRI method does not offer information about linear attenuation coefficient and this is one of basic challenges of this attenuation correction method [12]. The MRI attenuation correction procedure uses technique for image processing (segmentation and registration) for extracting the part of tissue structures. It is not possible to obtain information of all kinds of tissues, especially bone cortical [13]. The segmentation procedure divides the image into regions with uniform attenuation coefficient. Every region represents one kind of tissue according to the gray value.

When the image is divided into segments, the next step is determination of path way particle length from every pixel to the PET detector. Using this length and corresponding linear attenuation coefficient for energy of 511 KeV using equation (1) the attenuation correction factor is obtained. The attenuation correction is now applied for every pixel of the PET image.

C. Image registration

The image registration is matching of two or more images of the same object taken by the different imaging modality, different position or configuration. Using mathematical language, if we have two images $\mathcal{I}$ and $\mathcal{J}$ of the same object observed by the different modality method, the registration is founding the mapping function $\phi \circ \mathcal{I} \to \mathcal{J}$ which maps every pixel of $\mathcal{I}$ to $\mathcal{J}$. The difference $\phi \circ \mathcal{I} - \mathcal{J}$ represents displacement field of every pixel image $\mathcal{I}$. If matching process is represented as deformation the map function is then time varying function with normalized time t as the parameter. The velocity of deformation is given b:

$$\frac{d\phi(x,t)}{dt} = \mathrm{v}(\phi(x,t),t) \tag{8}$$

and $\phi(x,0) = x$ is satisfied.

The problem of finding corresponding map function is based on the minimization of functional

$$\int_0^1 \|\nabla \mathrm{v}\|^2 dt + \int_R \|\phi \circ \mathcal{I} - \mathcal{J}\| \cdot dR \tag{9}$$

where *R* is domain of image space [14]. The next step is applying the gradient-based optimization strategy for minimizing equation (9). Using obtained map it is possible to finish process of registration.

IV. RESULTS

In this study the images data set [15] is used. The study includes multimodal brain images of MRI, CT and PET.

In the Figure 3, the different imaging modality of the same brain cross-section is shown.

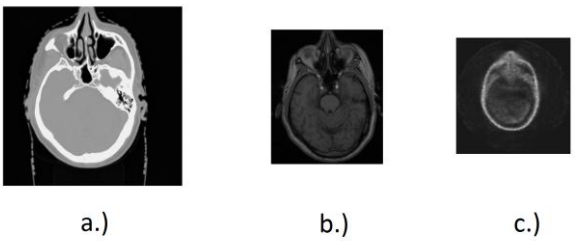


Figure 3. A Multimodal images of brain slices a.) CT image, b.) MRI image, c.) PET image

The first step in attenuation correction procedure is registration of the corresponding PET and CT images. The purpose of the registration is aligning the CT image with the PET image with the aim to map linear attenuation coefficient value in the space of the PET study observation. The different modalities have different values of the image resolution, (512x512 pixels for CT and 336x336 pixels for PET). The CT image has higher resolution and it is necessary to down sample this image by the factor 336/512. The spline interpolation has been employed.

The registration of the images represents procedure for finding an appropriate transformation that establishes correspondence between pixels of CT and PET images. In the Figure 4, the registration of the PET and CT images is shown.

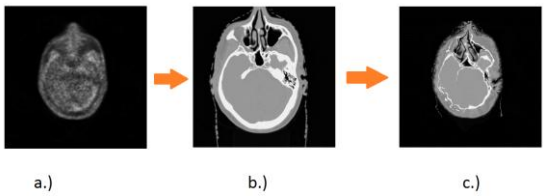


Figure 4. Registration procedure of the CT and PET image, a.) PET image with resolution 336x336 pixels, b.) Down sampled CT image, c.) Registered image.

The PET image is fixed and CT is warped by the corresponding transformation. When the images are matched, the linear attenuation coefficient from gray value can be calculated. It is assumed that maximum and minimum gray value corresponds to cortical bone tissue and the air respectively. The linear attenuation coefficient of the cortical bone for gamma photon energy of the 511 KeV is 0.1714 cm^{-1} and for air is 10^{-4} cm^{-1}. This value for the other tissue is assessment using interpolation between maximum and minimum value according to gray value. Using the calculated linear map the final PET image with attenuation correction is obtained, Figure 5 b.).

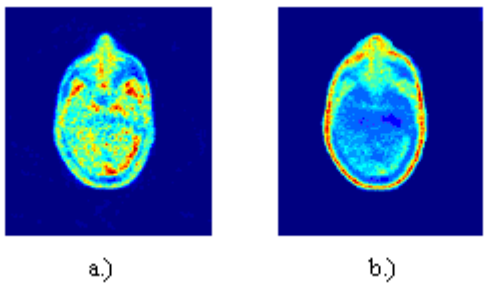


Figure 5. Attenuation correction a.) PET image before correction, b.) Corrected image

V. CONCLUSION

In this study the different imaging methods are presented. The advantage of multimodal imaging is obtaining a more quality information about tissue and organs due to fact that each of the imaging techniques has some limitations. The particular attention is given to the PET study and the problem of attenuation correction.

This problem can not be unambiguously solved due to uncertainty of the linear attenuation coefficient. One way to resolve this problem is using CT scanner image and combining it with PET. The CT scanner uses polyenergetic X-ray source and PET uses gamma photon with energy of 511 KeV. The main goal of attenuation correction is to convert this effective linear attenuation coefficient measured by the CT to attenuation coefficient for gamma photon energy. CT X-ray source generates photon energy in the spectra range. The attenuation coefficient represents average value for all energies from the spectra range due to high dependent between attenuation coefficient and energy of photon.

The approach in this study is based on linear conversion gray value of reconstructed image into linear attenuation coefficient with assumption that maximum and minimum values correspond to bone and air respectively.

ACKNOWLEDGMENT

This paper is part of project III41017 "Virtual human osteoarticular system and its application in preclinical and clinical practice", funded by the Ministry of Education and Science of Republic of Serbia. http://vihos.masfak.ni.ac.rs.

REFERENCES

[1] Fred A. Mettler, Jr., MD, MPH and Milton J. Guiberteau, MD, Essentials of Nuclear Medicine Imaging, 6th ed. Saunders; 2012.

[2] Hsieh, Jiang. "Computed tomography: principles, design, artifacts, and recent advances." Bellingham, WA: SPIE, 2009.

[3] Vadim Kuperman, Magnetic Resonance Imaging: Physical Principles and Applications, Academic Press 2000.

[4] Nikola Mijailovic, Jasna Radulovic, Aleksandar Peulic, Miroslav Trajanovic, Nikola Radulovic, CT SCANNER QUALITY ACCORDING TO EXPOSURE DOSE DURING SCANNING PROCEDURE, 8th International Quality Conference, Kragujevac, Serbia, 2014

[5] Radulović, J., Mijailović, N., Trajanović, M., Filipović, N., Radulović, N.,, Estimation of exposure dose of human head during CT scanning procedure using Monte Carlo simulation, 11th International Scientific Conference MMA 2012 - Advanced Production Technologies, Novi Sad, 2012, 20-21. September, pp. 513-516, ISBN 978-86-7892-429-3

[6] Mitchell, Jonathan, T. C. Chandrasekera, D. J. Holland, L. F. Gladden, and E. J. Fordham. "Magnetic resonance imaging in laboratory petrophysical core analysis." *Physics Reports* 526, no. 3 (2013): 165-225.

[7] Wagenknecht, Gudrun, Hans-Jürgen Kaiser, Felix M. Mottaghy, and Hans Herzog. "MRI for attenuation correction in PET: methods and challenges." *Magnetic Resonance Materials in Physics, Biology and Medicine* 26, no. 1 (2013): 99-113

[8] Zaidi H, Montandon ML, Alavi A. Advances in attenuation correction techniques in PET. PET Clin. 2007;2:191–217. doi: 10.1016/j.cpet.2007.12.002

[9] Soriano, A., A. González, A. Orero, L. Moliner, M. Carles, F. Sánchez, J. M. Benlloch, C. Correcher, V. Carrilero, and M. Seimetz. "Attenuation correction without transmission scan for the MAMMI breast PET." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 648 (2011): S75-S78.

[10] Burger, C., G. Goerres, S. Schoenes, A. Buck, A. Lonn, and G. Von Schulthess. "PET attenuation coefficients from CT images: experimental evaluation of the transformation of CT into PET 511-keV attenuation coefficients." *European journal of nuclear medicine and molecular imaging* 29, no. 7 (2002): 922-927.

[11] Kawaguchi H, Hirano Y, Yoshida E, Kershow J, Shiraishi T, Suga M, Ikoma Y, Obata T, Ito H and Yamaya T 2014 A proposal for PET/MRI attenuation correction with μ-values measured using a fixed-position radiation source and MRI segmentation Nucl. Instrum. Methods Phys. Res. A 734 156–61

[12] Steinberg, Jeffrey, Guang Jia, Steffen Sammet, Jun Zhang, Nathan Hall, and Michael V. Knopp. "Three-region MRI-based whole-body attenuation correction for automated PET reconstruction." *Nuclear medicine and biology* 37, no. 2 (2010): 227-235.

[13] Keereman, Vincent, Yves Fierens, Tom Broux, Yves De Deene, Max Lonneux, and Stefaan Vandenberghe. "MRI-based attenuation correction for PET/MRI using ultrashort echo time sequences." Journal of nuclear medicine 51, no. 5 (2010): 812-818.

[14] Avants, Brian B., Nick Tustison, and Gang Song. "Advanced normalization tools (ANTS)." Insight J (2009).

[15] http://www.osirix-viewer.com/dataset

DICOM Image Management Through Agents Based Systems

Dani Juliano Czelusniak[**], Erica Beatriz Fuscolim[**], Osiris Canciglieri Junior[*]

[*] Pontifical Catholic University of Paraná / Polytechnic School - Production and System Engineering Graduate Program (PUCPR/PPGEPS), Curitiba, Paraná, Brazil.

[**] Pontifical Catholic University of Parana, Polytechnic School, Production Engineering Undergraduate Program (PUCPR/EP), Curitiba, Paraná, Brazil.

dani.czelusniak@pucpr.br, ericafuscolim@hotmail.com, osiris.canciglieri@pucpr.br

Abstract— This paper presents a research that is developed inside "Products and Systems Design and Development" research group of Pontifical Catholic University of Parana (PUC-PR). In this research is studied, designed and developed a DICOM image loader with agents based system, that in a near future will be a important module of a dental prosthesis design decisions multi-agents software. In this paper, will be shown an overview of agent systems explaining its structure with a framework called JADE, used to provide agents software environment. The methods used to guide this research were, the bibliographical survey to list de concepts, and, states of art for the expert system project layout, develop and tests procedures. As a result, this paper shows that using agents' software techniques in expert systems development is a new way to manage data in a flexible way, offering forms to creating modular solutions that should have the capacity to deal with complex scenarios. It also allows, better software scalability and modularization for complex and specialized software solutions.

I. Introduction

Actually, is more evident in corporate environments, the application of software tools for automation and administrative control of processes in operational environment. The development and maintenance tools for software applications are in constant evolution to make the business processes development, more flexible and bug-free [10].

In this scenario, arises a new view for building applications, called agents software based systems [2]. This new architecture is gaining developers attention by the capacity to enable the applications development in a flexible form, because it permits that new software modules inserted in the agents environment can coexist with others, without necessity of modifying these agents "modules code", treated later as agent behaviors [3], [4].

Following, an introduction will be presented from the current business information technology context that aims for new challenges of software context with expert systems. Next, will be presented the JADE agents framework [1] and the advantages of its use before the conventional building software techniques.

This paper, therefore, presents how agent software behaviors need to be designed using agent-based systems techniques with JADE framework, to manage DICOM medical images portfolio. This acquaintanceship facilitates the constantly evolving of the developed system software [5]. To better understanding, this article presents this new software type, focusing on how agent software behaviors works, serving as a guide to develop agent's software to manage DICOM medical images portfolio, that is a module of a dental prosthesis design decisions multi-agentss system that is under development.

II. Agents Software Systems

Agents' software is a special kind of software that is work internally in autonomous ways, working well with the complexity of technological evolution. In high complex scenarios, is necessary use software, which has capacity of adaptation and provide mechanisms, which allows it, to take decisions in accordance with changes in their directives. This modality of information system commonly is known as "agents based systems" or "agents systems" [6], [7], including all kinds of agent, agents and multi-agents software. This way, agents systems are a software construction kind, which was born from Artificial Intelligence classes. These days, they are increasing distinguishability by the ability to permit applications development not only in modular, but modular and autonomous software modules.

Reviewing the literature about software development using artificial intelligence techniques to build agent based systems, seems that is the artificial intelligence is the study and creation of machines that displays human-like qualities, including ability to reasoning [11]. In this context, the agent software can maps the awareness in actions; their architecture varies by species and function of agent software, which can be a database, an intranet, or dedicated expert embedded software in some hardware. Its architecture is responsible for the agent's interaction. The relationship between software agents, intelligent software and architecture according with [7], shown in the Figure 1.

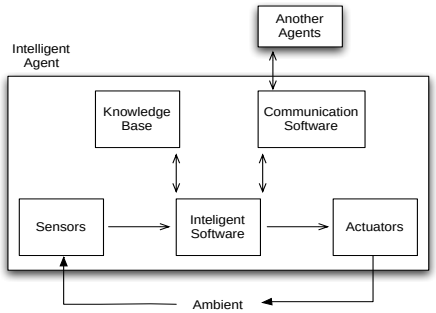


Figure 1 - Internal structure of agent software, adopted from [7]

Intelligent agents are software that have autonomous capacity to compare scenarios and execute actions, collect information, explore and learn in an environment according stimulus or detected perceptions, in the best way aiming execute it tasks faster and precisely mode as possible. Agent is anything that can be considered able to perceive their environment through sensors and act on this by actuators [7], for each possible perceptions, a rational agent must select an action that is expected to come to maximize their performance measure, given the evidence provided by a set of perceptions and by any internal knowledge of the agent.

JADE (Java Agent Development Environment) is one of the existing sets of software libraries, called by software developers as "framework", which is designed to develop applications with agents structures models [1]. Agents systems developed with this framework, will follow the standards defined by FIPA (Foundation for Intelligent Physical Agents). This is an IEEE (Institute of Electrical and Electronics Engineers) member organization, which arranges computational standards for agent, agents and multi-agents systems, guaranteeing interoperability between heterogeneous agents standards [6].

This framework is referenced in the software development tool that JADE is composed, by software elements that FIPA standards defined as necessary for their operation. This software elements set for agents software development, is called as agent platform by [1] and [2], with main reference architecture of FIPA [8] agent platform.

The conceptual JADE framework, to be presented by Figure 2, consists of agents and containers (that is a software environment where agents runs its lifecycle) the standard FIPA defines as necessary for its operation. All JADE Agent Platform has a special container, called the Main Container, which is the first container to be initialized when the application is loaded and. It contains two special agents and the communication service, which will provide functionality to the platform.

Conceptually, this framework was divided in some parts described below:

• AMS: System Management Agent, it oversees the access of other agents to the platform, and is responsible for control and authentication of agents' calls and the life cycle of services provided by the platform. Maintains in its structure a list with the identifiers of the agents along with their states.

• DF: Yellow Pages Service (or Directory Facilitator) is the service that provides the addressing mechanism of the agents, applying the metaphor of "yellow pages" of the agent platform.

• Message Transport Service (or ACC Agent Communication Channel), is the system component that controls the exchange of messages, managed the use of communication protocols.

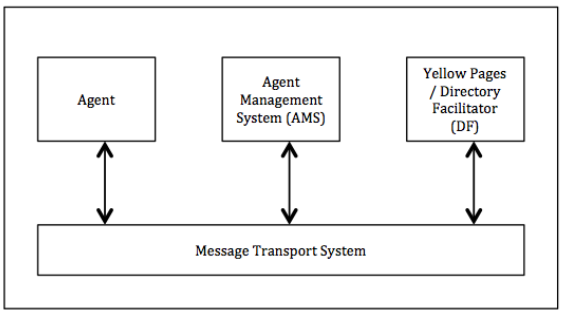


Figure 2 – Conceptual JADE Agent Platform, adopted from [1]

Following the specifications set by FIPA, the services of yellow pages and management agents communicate with the agents through language called FIPA-sl0. However, there may be used other communication language implementations from the FIPA standard.

Besides the basic elements needed for the operation of the JADE framework described above, below there is another figure that exemplify physical JADE container element. Each instance of the JADE environment, which is currently running, has at least one container, and agents residing therein. A set of assets containers is called Agent Platform, as shown in Figure 3. For example, in a computer network, instantiate a container computer and create mobile agents to flow through the network computers, seeking more interesting machine to perform certain processing.

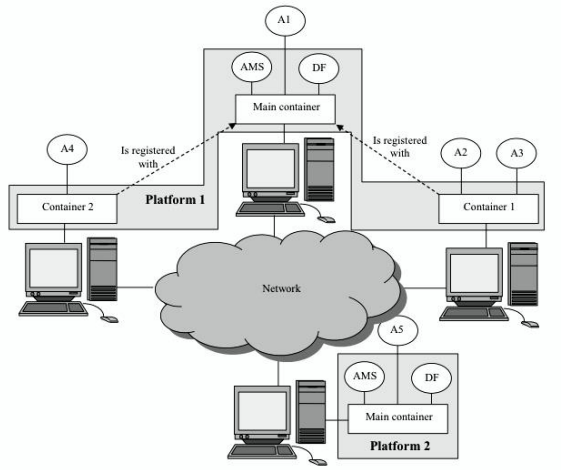


Figure 3 – Physical JADE Agent Platform, from [1]

Software tools with these characteristics have the possibility to make the analysis of large volumes of information, and can be executed and monitored, without requiring manuals tabbing and reading. For this, an agent software tool needs to be modular, to allow a shorter setting and adaptation to the desired environments, covering the concerns of professionals in short time periods; evolving to accompany the necessities, always guaranteeing best results.

According with [1] JADE provides a simplified interface for agents to access the AMS and DF services, simply by sending and receiving messages in standard-defined format. As these implementations are fully

standardized because they are very common, both classes can perform these tasks in a user interface, implemented through methods.

The Agent class in JADE, in the words of [1], is a *superclass*, which allows the software developer to instantiate its own agents. This implies that an agent inherits the characteristics, attributes and behaviors of its base class written in Java. In it, the actions are performed through behaviors, which are tasks that occur asynchronously and concurrently, booked by hidden internal mechanism to the software developer.

III. JADE Agent Software Structure

JADE Agents are instances of that perceive their environment through sensors and act when requested, being that its communication occurs through messaging [1], [6], [7], [10].

By codifying agents, it is necessary to inform the code of which class of JADE environment that the current agent is inheriting its basic structure [1]. In this article, the classes use the base class's Agent, by performance second policy inheritance the base class Agent, performing the second guidelines environmental policies, demonstrated below.

```
public class AgentSearch : Agent
```

In this context, seeking to understand the functioning of a system, its structure was divided into three distinct processes [1]:

Whenever a JADE agent is instantiated in a container, agent for acting Agent Controller automatically runs the *setup( )* method overload of the agent class, which is tasked to perform their initialization, as demonstrated by the following code fragment:

```
public override void setup()
```

During the execution of this method, operations that creating the agent description are performed so as to make your registration on Directory Facilitator. After these procedures, are initialized the behaviors of agents, through the add Behavior method, according in the following example, where it is added the behavior of receiving messages.

```
addBehaviour(new ReceiveMessageFromMControl(args));
```

After initiating the behaviors, they are executed and after finalization the method that performs the agent is executed. In the same form as the *setup ( )*, It is also necessary to put your overload into effect, and this method is called take down, as exemplified in the sequence:

```
public override void takeDown()
{
   FileAgent.Log("Agent" + getLocalName() + " has been
finished...");
   this.takeDown();
}
```

IV. JADE Agent Software Behaviors

The agents behaviors are actions that they have execute in face of a perceived stimulus in the environment [6], [7], as an incoming message, for example. Can be initialized from the *setup method ( )* as already discussed agent, but they can also be initialized at any moment from other behavior [1]

In its structure, basically reside two implementations for the behavior to perform actions. The first is by the method *action( )*, which owns the code that will be executed when the behavior is activated. The second implementation is responsible for finalizing the behavior and is implemented by the method *done( )*, It is named after the implementation of behavior, be it for having achieved an objective or by finalization of itself as defined implementation by type of behavior, described in the sequence.

"One Shoot" Behavior: The simplest type is implemented in a behavior. The method *action( )* the agent behavior is executed only once and at the end, the code is finalized directly. It, does not appear the method *done( )* as will be described in other behaviors that this always returns the logical value true by the end of its execution.

```
public class Name_Behavior : OneShotBehaviour
{
   public override void action()
   {
      // Do α
   }
}
```

"Ciclic" Behavior: This behavior never finalizes its execution, executing the action whenever it is activated. As opposed the previous type of behavior, this behavior always returns a logical value false.

```
public class Name_Behavior : Ciclic
{
   public override void action()
   {
      // Do β
   }
}
```

"Simple Behavior" Behavior: Is the behavior commonly used in systems development agents with the platform JADE. Have the possibility to perform different operations depending on your state, or the way it is activated; according demonstrates the code fragment below. It can be observed the overdid method *action( )* (1), and tests of conditions for execution of tasks in (2) e (3) and finalization of behavior in(4) with the execution of the method *done( )*.

```
public class ReceiveReplies : SimpleBehaviour
{
(1) public override void action(){
(2) if (conditionA == "value"){
       // Do γ
    }
    else{
(3)    if (conditionB == "value"){
            // Do δ
       }
    }
}
(4)  public override bool done(){
        return completed;
     }
}
```

V. JADE Agent Software Messaging

Messages in JADE follow the standards and rules set by FIPA, available at www.fipa.org. The message exchange mechanism is executed a synchronously between the agents. Each agent has a kind of "p.o. box" where the messages are received and every receiving message is notified to the agent, according to what presents the Figure 4 constructed from [1] that presents the form through which is effected the exchange of messages among the agents.

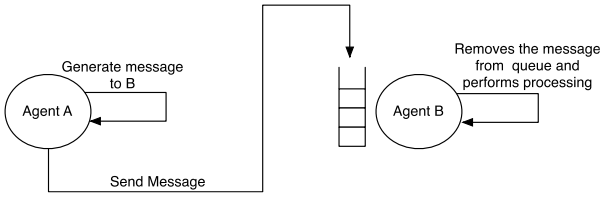


Figure 4 – Physical JADE Agent Platform, from [1]

The code fragment demonstrated below presents the form through which a message is mounted and sent. Initially, a message object is created (1) from the class, which has the structure that enables the handling message objects. Following are inserted in the message the receiver (2), the sender (3) and the content (4). After that, the message is sent (5) from the agent to your receiver.

```
(1)ACLMessage myMsg = new ACLMessage(ACLMessage.INFORM);
(2)myMsg.addReceiver(destination);
(3)myMsg.setSender(sender);
(4)myMsg.setContent(content);
(5)myAgent.send(myMessage);
```

For the agent to be able to put into effect the receiving of messages, it is also necessary to create a class-based Message Template method, as described in (1). So, is effected the receipt of the message in (2). In the case of system developed, one of the agents there which is the sender of the message to be able to realize the treatment of the data according can be seen in (3).

There is also the possibility to perform the processing of the personal data, according to the content of the message that was received by the agent, triggering another behavior. It is possible to be open the content of a message via the method *getContent( )*.

```
(1)MessageTemplate receivedMsg = MessageTemplate.MatchPerformative(ACLMessage.INFORM);
(2) ACLMessage myMessage = myAgent.receive(receivedMsg);
(3) if (aCLMessageReplay.getSender().getLocalName() == "sender"){
    // Performs the processing of the message
}
```

VI. Case Study: Design of an Agent Software With a Behavior to Load DICOM Images Data

Structuring an Expert Agent System Prototype to Support Dental Prosthesis Design Decisions. The prototype of an expert agent system to support dental prosthesis design decisions will consists of three autonomous agents, that will can exchange messages one with other and understand concepts in an ontology described in OWL format organized with Protege Ontology software. These agents will be denominated as Agent Master Control, Agent List File and Agent Search. This model of agents based software was worked firstly in [12] and updated in [13].

Specifically, the "Agent List File" will be responsible for read and load the files recorded in a directory. This files are generated by tomography and magnetic resonance equipment's, in the DICOM format, making possible the image processing according to the necessity, extracting characteristics or information that will be translated or shared with other systems. The DICOM standard, adopted from 1993, made the image processing treatment by algorithms since the information obtained directly from the hardware, to support health professionals [14], [15], [16].

These agents that will be developed in this first version of the expert system software are classified as "simple reactive agents", by the fact that the agents choose their actions on the basis of on his present understanding of the fact, derived from environmental stimulus, discarding operations coming from their historical frame of reference. There are plans to update the agent's behaviors of the next version of this expert system to a level of basis reasoning.

On agents' expert system loads, the main container that has agents is instantiated. Then, a control agent of proper JADE present in the framework so called Agent Controller. It instantiates all the agents in the container and activates them, by its *start( )* the container method. Once the method is instantiated and activated, the agent runs your own *setup( )* behavior and begin work.

Figure 5 illustrates the inner working of expert system agent software prototype. After the agents initiating, the "Agent Master Controller" request to the "Agent List File" to list the files that are in directory intended for DICOM files. This agent will returns to the "Agent Master Controller" a list with all files founded. Receiving this list, the "Agent Master Controller" opens each file, read it, extracts necessary data, performs its conversion to a format that will be used by the expert system and sends them, to the "Agent Analyzer". Receiving this information, the "Agent Analyzer", performs the information analyzing by requested criteria and returns it result to the "Agent Master Controller". After, the agents based expert system processing is completed.

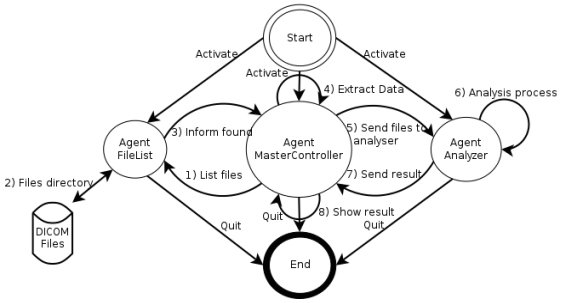


Figure 5 - Structural Diagram of Agent Expert System Prototype adopted from [12] and [13].

The agent that will be the sponsor by the analysis, the "Analyzer Agent", to perform it work will need data stored in ontology format. To improve this analysis behavior in agent software will be designed an ontology that provides patterns to the software agent analyzer behavior. With these patterns, the agent will have a prosthesis details collection, and will be able to "understand" different kinds of dental prosthesis parts needed for each situation.

VII. Conclusions and Results

This article has presented a research made to know the viability of develop an agents expert system with ontologies, to support DICOM image load and future analysis by agents based software. This research is a portion of a major Pontifical University of Parana, which is under design and development by Products and Systems Design and Development research group. This project that aims to make dental images treatment, to support design decisions in dental prosthesis processes. It shows that is incontestable the importance of information systems to support any kind of professionals during decision making processes. This need becomes more evident when have risk of errors that can seriously compromise human health or esthetics. Nevertheless, specifically in process of images portfolio management, like related DICOM format, where it becomes the need of rendering aspects that are not phrase. Will make necessary the interpretation of the graphical information with highly specialized skills of professionals.

About the agent software, its source code is presently under development. In first tests at the laboratory, shown that different agents can analyze the same DICOM images set at the same time, without an agent causes disturbs in the search and analysis process of another agent are running in an asynchronous way.

Under this scenario, aims that the agent systems can better assist in the correct classification of information precisely by the fact of effectively process large volumes of data by its characteristics of work in autonomous, modularized and associative way. Each agent of software is unique, but works by communicating (asynchronously, by predefined communication protocol) with the others, extracting information as request, a second agent, that passes this information to be treated by one third agent that can understand this kind of data, and so forth. This process might be concurrent and the professional with his experience skills will be able to choose the best result or even make a composition thereof.

About the JADE framework, is an alternative for developing agents based software, to meet FIPA specifications, guaranteeing interoperability with other systems which operating under this standardization, achieving a greater level of integration. In this sense, can be perceived that the ability of adaptation, reuse inherent in agents systems, in conjunction with the use of development and tools like JADE makes possible the expert software applications development which can be executed in most operating systems, allow for the development of software tools applied to the actual dental prosthesis professionals requirements. Is estimated that this new conceptions for expert software applications development, allied to the evolution of new technologies can build knowledge bases and practices, to feed processes of expert software developing with more efficient and flexible models.

References

[1] F. Bellifemine, G. Caire, T. Trucco, G. Rimassa. JADE Programmers Guide. Available at Jade Developers SVN Service at <https://avalon.cselt.it/svn/JADE/trunk> (2012).

[2] J. M. Corchado, R. Laza, L. Borrajo; J. C. Yañez, M. Valiño. Increasing the Autonomy of Deliberative Agents with a Case-Based Reasoning System. International Journal of Computational Intelligence and Applications. V 3, N 1 (2003).

[3] D. J. Czelusniak. Agents software architecture based for of information sources portfolio management on the web. Doctoral Thesis in Post Graduate Program in Production Engineering and Systems. Federal University of Santa Catarina. Florianópolis, Santa Catarina, Brazil (2013).

[4] D. J. Czelusniak. Agents system proposal to support skills management. Master Dissertation in Production Engineering. Federal Technology University of Paraná. Paraná, Brazil (2007).

[5] A. C. B. Garcia, J. S. Sichman. Agentes e Multiagentes, in: Sistemas Inteligentes: Fundamentos e aplicações. Organização Solange Oliveira Rezende. Barueri, SP. Editora Manole (2005).

[6] E. Rich, K. Knight. Artificial Intelligence. Ed. McGraw-Hill. International Edition. Singapure (1991).

[7] S. Russel, P. Norvig. Inteligência Artificial. 2ª. Edição. Editora Elsevier. Rio de Janeiro, RJ (2004).

[8] The Foundation for Intelligent Physical Agents. FIPA. Information at: <http://www.fipa.org/>

[9] M. Greaves. Semantic Web 2.0. IEEE. Intelligent Systems (2007).

[10] S. Liao. Technology management methodologies and applications: A literature review from 1995 to 2003. Elsevier. Technovation 25 (2005).

[11] A. C. B. Garcia, J. S. Sichman. Agentes e Multiagentes, in: Sistemas Inteligentes: Fundamentos e aplicações. Organização Solange Oliveira Rezende. Barueri, SP. Editora Manole (2005).

[12] D. J. Czelusniak. Agents' software architecture based for information sources portfolio management on the web. Doctoral Thesis in Post Graduate Program in Production Engineering and Systems. Federal University of Santa Catarina. Florianópolis, Santa Catarina, Brazil (2013).

[13] D. J. Czelusniak. Agents' system proposal to support skills management. Master Dissertation in Production Engineering. Federal Technology University of Paraná. Paraná, Brazil (2007).

[14] A. L. Szejka, M Rudek, O. Canciglieri Junior. A Reasoning System to Support the Dental Implant Planning Process. 19th ISPE International Conference on Concurrent Engineering. Trier, Germany (2012).

[15] D. Grauer, L.S.H. Cevidanes, W.R. Proffit. Working with DICOM craniofacial images. American journal of orthodontics and dentofacial orthopedics: official publication of the American Association of Orthodontists 136 (2009).

[16] R. N. J. Graham, R.W. Perriss, Scarsbrook AF DICOM demystified: A review of digital file formats and their use in radiological practice (2005).

Development of Web-available Models of Human Spinal Vertebrae for Biomedical Engineering Research and Education

Milan Blagojević *, Miroslav Zivković *
* University of Kragujevac, Faculty of Engineering, Sestre Janjić 6, Kragujevac, Serbia
blagoje@kg.ac.rs, zile@kg.ac.rs

***Abstract* - Sharing different types of models over a long distance using Internet has great potential as a resource for research and education. This paper presents the development of different 3D model types of human spine and web-based application for user access and interactive exploration. The remote visualization of models is powered by O3D plug-in/WebGL, GLview 3D Plugin, and STLDroid. End-user is able to review online in real time, and download high-definition 3D models of spinal vertebrae. The presented solution could improve education in the field of biomedical engineering and become useful tool in spinal research.**

I. INTRODUCTION

The Internet is strongly influencing all aspects of present-day life including learning [1-4]. Many examples of web-based applications in engineering education are described in the literature [5-8]. Web-based training offer an elegant solution to the need for better training, since realistic and configurable training environments can be created [6, 7]. This can bridge the gap between basic training and performing the actual intervention on patients, without any restriction for repetitive training [7].

Today's Internet knows many web-based solutions in this area. The Body Browser [9] from Google, ZygoteBody [10] from Zygote, and BioDigital Human [11] from BioDigital Systems offer Internet users a new, hi-tech way to explore the human body. These online resources obviously can help students at every level. These new tools promote knowledge and learning while having fun. However, models of organs on these applications are not publicly available.

Many sharing platforms are being currently developed thanks to users' contribution. The ability to share their work and ideas with a large audience of users is an incentive to create and add new content. The VAKHUM project [12] has developed an interactive database of human organs for educational, research and industrial purposes. Users can access the database through a virtual interface and download high-quality data for their own applications, or take an online class on functional anatomy. The 3DVIA [13] empowers anyone, at any skill level, to create and publish professional quality, lifelike 3D applications and experiences, by providing an integrated suite of 3D software and authoring tools, a large marketplace of 3D models and 3D content and a growing community of 3D professionals. GrabCAD [14] also provides different types of free CAD models. All models are created by GrabCAD members and shared for free. These models are general purpose and were obtained without CAD modeling information on how they are faithful to realistic models.

Over the years, finite element method (FEM) has established as appropriate for the predicting of the biomechanical behavior [15-17]. The results of the finite element models may be trusted if they take into account the actual geometry of the domain and realistic boundary conditions [18, 19].

Recent improvements in 3D scanning technology allow a reliably and accurately digitizing of the external shape of many physical objects with high definition and accuracy [20-23]. However, many researchers do not have access to scanning facilities, dense polygonal or quality finite elements models.

In this paper we present the development of a web-based application for user access and interactive exploration of three-dimensional biomedical models through an intuitive interface. It is presented the whole pipeline from the creation of a high resolution 3D and FEM models of each human spine vertebra to its remote rendering on user's computer without any loss of details or accuracy, ensuring enjoyable learning with good academic quality and flexibility, as well as quality resource for R&D.

II. METHODS

A. *Development of Human Spine 3D Models*

In human anatomy, the vertebral column (backbone or spine) is a column usually consisting of 24 articulating vertebrae, and 9 fused vertebrae in the sacrum and the coccyx. It houses and protects the spinal cord in its spinal canal and allows complex motions while providing stability and protection for the spinal cord during a variety of loading conditions. The observed spine was found to be free from spinal disease and trauma (Fig. 1).

We have adopted a direct digitizing approach that provides an alternative method to common method (by stacking coronal and sagittal computed tomography (CT) images) in capturing the highly irregular bony structure of spine. Accordingly, this approach has allowed a better mesh representation of the geometry, which is of the essence in the study of stress distribution patterns in the spine. For each vertebra of spine, overall process consists of the following phases: (a) geometric model development, (b) volumetric model development, (c)

surface reconstruction, (d) FEM model development, and (e) upload to web server. After all processing steps, high definition and high density surface models are obtained.

Figure 1. Spine to be digitized

B. Digitizing

3D scanners accurately scan and capture the surface of objects and provide real 3D data [24-26]. Scanning of 3D shape of each spinal vertebra was performed by using optical measuring system GOM ATOS IIe [27-29]. Overall scanning process with the ATOS IIe system covers following phases: (a) calibration, (b) preparation and setting of device, (c) preparation and setting of measurement object, (d) measurement/scanning, (e) processing of measured/scanned data, and (f) post-processing (processing of results).

Figure 2 depicts some phases in the scanning procedure. Figure 2a shows the vertebra being digitized in the vice while the ATOS IIe was used to extract the geometrical data by moving its sensor across the vertebra. Figures 2b and 2c show the scanning vertebra process. Complex geometry is scanned in two measuring project: first project deals with the top half of the model (Fig. 2b), and second project deals with the bottom half of the model (Fig. 2c). Measurement projects are connected through common uncoded reference points that are available (visible) in both projects. Structured-light 3D scanners projects white light fringe patterns on an object surface. The deformations of the pattern are captured by two measurement cameras (Fig. 2d).

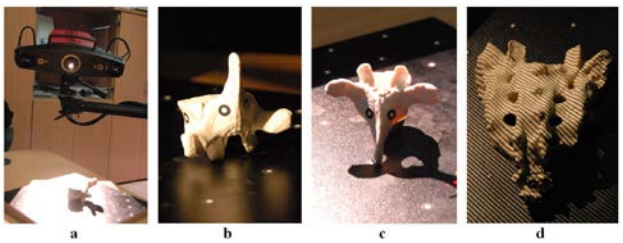


Figure 2. Digitization of specific vertebrae

Complex geometries of the measuring object are digitized through sufficient number of individual measurements. In order to digitize specific part of free-form 3D surface, at least one measurement is required. For each spinal vertebra, more than sixty individual measurements were acquired due to the topological complexity of the physical model, moving the scanner in different position around the object. Figure 3 shows the different stages in generating points on surface of volumetric models. Configurations shown in figure 3 correspond to point cloud generated after 2nd (a), 11th (b), 18th (c), 32th (d), 48th (e), and 62th (f) individual measurement. A dense point cloud is then produced through software. The point cloud represents the digital model of the scanned object. After the poligonization process, the result is a detailed triangulated mesh. The obtained models are registered into the appropriate coordinate system.

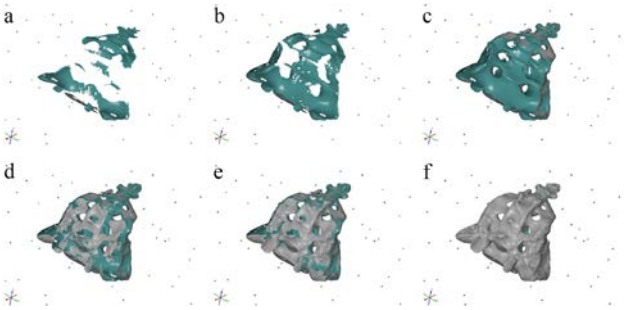


Figure 3. Acquisition of sacrum's 3D shape of scanned object using ATOS IIe system

C. Models' development

Further processing (erasing, reparation, relaxation, decimation, detection of contour lines…) of triangulated mesh is conducted by using Polygon Phase in Geomagic Studio. The key step in development of high quality surface is decomposition of polygonal models in patches or the number of patches covered by grid (the union of Bezier curves). Creating of the model's surface is carried out with introduction of certain assumptions and approximations. Surface reconstruction is performed using generative shape design (CATIA V5) module. Process of reconstruction is ended by export of surface model. Furthermore, the surface created was compared with the actual (digitized) vertebra to ensure its geometric integrity. Following this procedure, volume modeling was thus developed for the entire spine structure. Figure 4 shows the volume rendering for randomly selected spinal vertebrae: C1 (Atlas), L2, T9, and S1-S5+Sacrum.

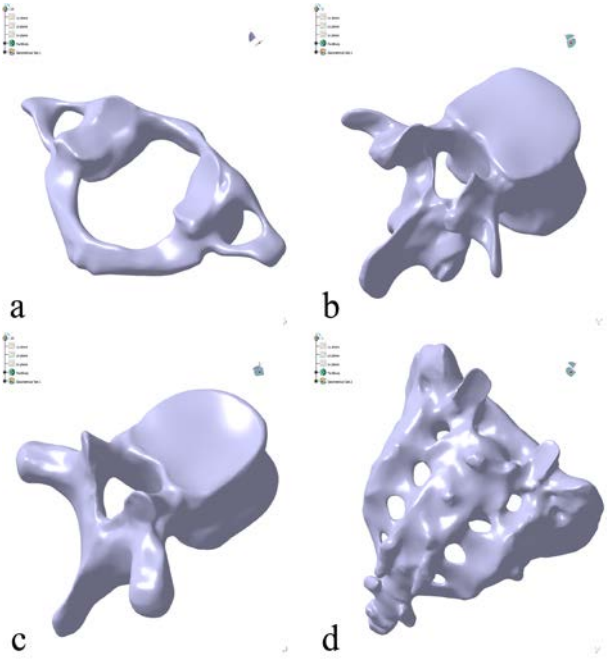


Figure 4. CAD models of random vertebrae: (a) C1 (Atlas), (b) L2, (c) T9, and (d) S1-S5+Sacrum

D. Interactive Web Resource Development

The project uses the open source code JOOMLA to implement a repository of 3D models. In terms of system architecture, the portal is created with Apache as a server and PHP as scripting language for dynamic pages. The data and metadata related to the models, as well as a hierarchical organization of the model parts and the link between model parts and relative information is stored in a full open-source object-relational MySQL. Operations delegated to the client are limited to decompression of the records and keyboard and mouse input management.

Raw COLLADA files [30] exported from Meshlab are converted by the COLLADA Converter for use by the O3D JavaScript API. The volumetric models are stored in *.o3dtgz file, which supports binary compression, useful for sending lighter data over the Internet, and hypertext mark-up language. The remote visualization of models, stored in *.o3dtgz file, is powered by O3D plug-in [31], as the most appropriate solution for developed viewer. The O3D JavaScript application code is completely contained in an HTML document that is loaded into a web browser. O3D software communicates with system's graphics hardware through either the OpenGL or Direct3D library. The framework supports known HTML interaction standards and event models on 3D objects. Nowadays, JavaScript implementation of the O3D API using WebGL replaces the original O3D plug-in. To be able to enjoy the power and the knowledge of this new Google asset, users must have a WebGL capable Web browser.

GLview 3D Plug-in [32-34] is a free 3D viewing component enabling presentation of finite element analysis data generated by GLview Inova or the GLview Express Writer. GLview Inova support standard simulation codes (PAK, ABAQUS, ANSYS, FEMAP, I-DEAS, LS-DYNA, MSC. Marc, MSC. Nastran, PAMCRASH, RADIOSS, CGNS, Fluent). The analysis results can subsequently be distributed in a compressed file by utilizing the free tools GLview Express and GLview 3D Plug-in. The plug-in is embedded into web environment and enables full 3D interactivity and high performance graphics. GLview 3D Plugin reads encrypted VTF and VTFx files, that can be created using GLview Inova or wrote as a solution calculated by software PAK [16, 17] using the GLview Express Writer. The files can contain the 3D model, selected scalar and/or vector results, display and animation settings, feature extractions (iso-surfaces, cut planes, particle traces) and annotations.

Due to the rapid development of mobile platforms and operating systems, next step was to enable the developed models are displayed on these devices. With this kind of technology the user can utilize hardware tools (netbook, PDA or smartphone), even despite the limited bandwidth connection, to access the platform. For the Android platform, there is a number of open source software that can display generated models. Within this part of project the open source STLdroid viewer was implemented. Using dedicated links, models stored on the dedicated server can be interactively displayed on users' devices. These models are smaller and have a smaller number of details due to still limited performances of these devices.

III. RESULTS AND DISCUSSION

Application can be accessed through the Internet. It requires that the users register once to the system database. The platform allows efficient and effective sharing of high quality 3D models without the need for the end user to download it to the local PC.

The Figure 5 shows a block diagram of the developed application. The user, through the dedicated web page, has to run O3D and/or GLview plug-in which installs and automatically configures the client to access the graphics application. It is not required to the user to know any configuration features or install other kind of applications. Once the plug-in is activated, navigation in the view is done using the mouse.

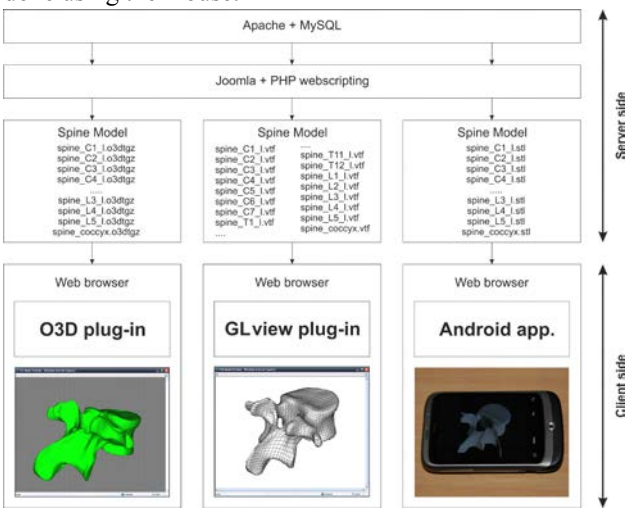


Figure 5. Model implementations in web-based application

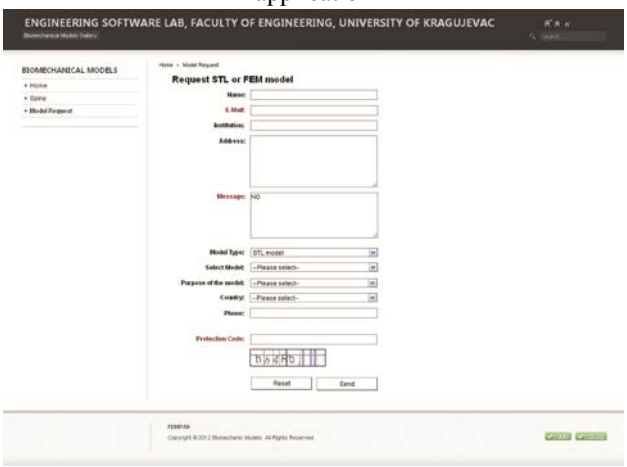


Figure 6. Model Request form in web browsers

At this stage, models are uploaded exclusively by paper authors, due to registration and verification of the content. All the models created so far completely free accessible to registered visitors of the web portal [35], figure 6. Users are quested to send basic personal information, model type (STL or FEM model), select one of vertebras, and define purpose of downloaded model. Improvements and FEM models tests are made continuously.

IV. CONCLUSION

This paper describes the technique and results in development of web-based application providing end-user with intuitive navigation interface for exploring models developed using CAD/CAM/CAE, as well as section for download of developed models.

The goals of developed web-based application is to: (a) bridge (mostly technical) gaps between biologists, doctors, student and engineers by providing knowledge and training in finite element modeling and analysis of complex biological and biomechanical structures, (b) provide resources to the community in the form of geometry, polygonal and finite element models, (c) promote the integration of biology and engineering by facilitating research and education.

The following types of models are available: solid model, volumetric watertight model, and finite element mesh. Acquisited point clouds and models present virtual shape of each vertebra identical to real one. Improvements and FEM models tests are made continuously. The software characterizes efficient knowledge transfer for spinal vertebrae using 3D-realtime display and animation, bringing good visibility and students' satisfaction. Presented objects become personal; they can individually be zoomed, rotated and animated. The user has the specific artifact virtual in his hand and can explore the product interactive and personalized.

ACKNOWLEDGMENT

This work is a part of ongoing research on "Software Development for Coupled Multiphysics Problems" at Laboratory for Engineering Software (University of Kragujevac, Faculty of Engineering, Kragujevac, Serbia). The research is supported by the Serbian Ministry of Education and Science under the grant TR-32036.

REFERENCES

[1] R. N. Coger and H. V. De Silva, An Integrated Approach to Teaching Biotechnology and Bioengineering to an Interdisciplinary Audience, Int J Engng Ed 15 (1999), 256-264.

[2] L. Senhadji, M. Siebes, J. V. Sloten, and N. Saranummi, Biomedical Engineering Trends in Europe, IEEE Eng Med Biol Mag 23 (2007), 12-13.

[3] T. C. Pilkington, F. M. Long, R. Plonsey, J. G. Webster, and W. Welkowitz, Status and trends in biomedical engineering education, Eng Med Biol Mag 8 (1989), 9-17.

[4] T. R. Harris, Recent advances and directions in biomedical engineering education, Eng Med Biol Mag IEEE 22 (2002), 30-31.

[5] A. Guarnieri, F. Pirotti, A. Vettore, Cultural heritage interactive 3D models on the web: An approach using open source and free software, Journal of Cultural Heritage 11 (2010), 350–353.

[6] I. Song, J. Yang, A scene graph based visualization method for representing continuous simulation data, Computers in Industry 62 (2011), 301–310.

[7] M. Cote, J. A. Boulay, B. Ozell, H. Labelle, and C. E. C. Aubin, Virtual reality simulator for scoliosis surgery training: Transatlantic collaborative tests, Haptic Audio visual Environments and Games 2008.HAVE 2008.IEEE p 1-6. 10.1109/HAVE.2008.4685289.

[8] Y. Yuan, L. Qi, S. Luo, The reconstruction and application of virtual Chinese human female, Computer Methods and Programs in Biomedicine 92 (2008), 249–256.

[9] http://bodybrowser.googlelabs.com/

[10] http://www.zygotebody.com/

[11] http://www.biodigitalhuman.com/

[12] http://www.ulb.ac.be/project/vakhum/

[13] http://www.3DVIA.com

[14] http://grabcad.com/library

[15] K. J. Bathe, Finite Element Procedures in Engineering Analysis, Prentice-Hall, 1982.

[16] M. Kojić, R. Slavković, M. Živković and N. Grujović , Finite element method I – Linear analysis, Faculty of Mechanical Engineering in Kragujevac, University of Kragujevac, Kragujevac, 1998.

[17] http://mfkg.kg.ac.rs/fempak/

[18] S.S. Hu, C. B. Tribus, R. K-B. Tay and N. N. Bhatia, Scoliosis section. Disorders, diseases and injuries of the spine. In: H. B. Skinner, Ed., Current Diagnosis and Treatment in Orthopedics, 4th edition, chap. 5, Lange edical/McGraw-Hill, New York, 2006, pp 255-269.

[19] L. Y. Griffin, Ed., Scoliosis. In: Essentials of musculoskeletal care, 3rd edition. American Academy of Orthopaedic Surgeons, Rosemont, IL, 2006, pp 928-931.

[20] P. Patias, E. Stylianidis, M. Pateraki, Y. Chrysanthou, C. Contozis, and T. Zavitsanakis, 3D digital photogrammetric reconstructions for scoliosis screening, Commission V, WG V/6, Proceedings of the ISPRS Commission V Symposium, Image engineering and vision metrology, Dresden, Germany, 25-27 September 2006, pp 1682-1750.

[21] F. Berryman, P. Pynsent, J. Fairbank, and S. Disney, A new system for measuring three-dimensional back shape in scoliosis, Eur Spine J 17 (2008), 663-672.

[22] D. S. Shina, K. Leea, D. Kim, Biomechanical study of lumbar spine with dynamic stabilization device using finite element method, Computer-Aided Design 39 (2007), 559–567.

[23] H. L. Mitchell, I. Newton, Medical photogrammetric measurement: overview and prospects, ISPRS Journal of Photogrammetry & Remote Sensing 56 (2002), 286– 294.

[24] C. Rocchini, P. Cignoni, C. Montani, P. Pingi and R. Scopigno, A low cost 3D scanner based on structured light, EUROGRAPHICS 2001, Volume 20 (2001).

[25] D. Lanman and G. Taubin, Build Your Own 3D Scanner: Optical Triangulation for Beginners, SIGGRAPH 2009 and SIGGRAPH Asia 2009 Courses.

[26] M. Živković, M. Blagojević, D. Rakić, The annual report on the use of received and installed capital equipment: 3D Digitization Systems ATOS IIe and TRITOP, Ministry of Science and Technological Development of the Republic of Serbia, Faculty of Mechanical Engineering in Kragujevac, Kragujevac, 2007, 2008, 2009, 2010.

[27] ATOS User Information, ATOS IIe and ATOS IIe SO (as of Rev. 01) Hardware, GOM mbH, 2008, Braunschweig, Germany.

[28] ATOS User Manual Software, ATOS v6.01, GOM mbH, 2008, Braunschweig, Germany.

[29] M. Blagojević, The Application of Optical Measuring Systems in Modeling and Simulation, Diploma work, Faculty of Mechanical Engineering in Kragujevac, Kragujevac, 2009.

[30] M. Barnes and E. L. Finch, COLLADA – Digital Asset Schema Release 1.5.0, Specification, April 2008.

[31] http://code.google.com/p/o3d/

[32] T. Alstad, Post processing and reporting from FEA simulations, NST Users Conference, September 2010.

[33] C. Muthanna, T. H. Hansen, B. Pettersen, and H. Holm, International Conference on Computational Methods in Marine Engineering MARINE 2009, Efficient Visualization of the Flow Field Behind A Grid of Circular Cylinders Using GLview: A Comparative Study Between Numerical and PIV Experimental Studies.

[34] T. Alstad, T. H. Hansen, Combined Visualization, NAFEMS Nordic, April 2009.

[35] http://mfkg.kg.ac.rs/fempak/bioeng/index.php?option=com_smartformer&Itemid=49

Fuzzy Ordering Implementation Applied in Fuzzy XQuery

Supaporn Kansomkeat*, Sukgamon Sukpisit*, Apirada Thadadech*,
Pannipa Sae Ueng** and Srdjan Skrbic**
* Prince of Songkla University, Department of Computer Science, Songkhla, Thailand
** University of Novi Sad, Department of Mathematics and Informatics, Novi Sad, Serbia
supaporn.k@psu.ac.th, sukgamon.s@psu.ac.th, apirada.t@psu.ac.th,
pannipa@dmi.uns.ac.rs, srdjan.skrbic@dmi.uns.ac.rs

***Abstract* — Fuzzy XQuery is the extension of standard XQuery language that allows fuzzy values in the query condition statements. Relational operators are not only required and possible in crisp value cases but also for fuzzy values. When relational operators are included in the query, it is necessary to provide means for comparison between fuzzy sets. These fuzzy relational operators are typically used in two fuzzy sets comparison case, but can also be used with some aggregate functions like MIN, MAX, and SUM. The aim of this paper is to present the algorithms for the implementation of fuzzy relational operators. Our algorithms compare the horizontal positions of two fuzzy sets and calculate the ordering value based on partial fuzzy ordering proposed by Bodenhofer. Moreover, we developed a GUI application and evaluated our approach with 360 fuzzy ordering cases. The experimental results show that our algorithms are capable of calculating fuzzy ordering values with various types of fuzzy values correctly.**

I. INTRODUCTION

Recently, fuzzy extensions are proposed to handle vague, ambiguous, uncertain, imprecise or incomplete information. Campi et al. [1] introduced fuzzy extensions to XPath named FuzzyXPath that used to query XML data based on the fuzzy set theory. Fredrick and Radhamani [2] introduced fuzzy XQuery to retrieve data from native XML database. Skrbic et al. [3] introduced PFSQL (Prioritized Fuzzy Structured Query Language), which is an extension of SQL (Structured Query Language). PFSQL uses the prioritized fuzzy logic to retrieve data from a fuzzy relational database. In 2012, Ueng and Skrbic [4] proposed fuzzy extensions to standard XQuery. Their query system retrieves data from native XML database based on prioritized fuzzy logic. In 2014, they implemented an interpreter for fuzzy XQuery in their project called FXI (Fuzzy XQuery Interpreter). Users can query data with priority and threshold keywords in the condition statement and define fuzzy values used as search conditions in the query.

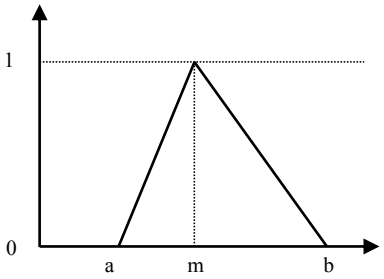
Including fuzzy relational operators in FXI is a very promising idea. In this way, fuzzy XQuery queries would be able to provide flexible comparisons between fuzzy sets that represent vague data. Relational operators on fuzzy sets are binary operators, which are able to compare two fuzzy sets: $<, \leq, \geq$ and $>$. Furthermore, fuzzy relational operators can be used with some aggregate functions like MIN, MAX, and SUM. In this paper, we propose a method to calculate fuzzy relational operations between two fuzzy sets and give its implementation. The proposed method is general and may be used with different types of problems. For example, it can be applied to fuzzy XQuery or PFSQL.

This paper is organized as follows. In the next section, we introduce algorithms for fuzzy relational operator calculations. Our implementation and testing results are presented in Sections 3 and 4, respectively. Section 5 is the conclusion.

II. FUZZY ORDERING CALCULATIONS

A. Membership functions

There are five different types of fuzzy membership functions used in [4]: triangle fuzzy number, trapezoidal fuzzy number, interval, fuzzy shoulder and crisp value. Figure 1 shows the shape of a fuzzy triangle membership function.



$$\mu_A(x) = \begin{cases} 0, & \forall x, x \leq a \wedge x \geq b \\ \dfrac{x-a}{m-a}, & \forall x \in (a, m] \\ \dfrac{b-x}{b-m}, & \forall x \in (m, b) \end{cases}$$

Figure 1 Fuzzy triangle number and its membership function

Definition 1 A fuzzy set A over universe X is determined by its characteristic (membership) function [5],

$$\mu_A: x \to [0,1],$$

where, for every $x \in X, \mu_A(x)$ is interpreted as membership degree of element x to fuzzy set A. Value $\mu_A(x) = 0$ denotes that element x does not belong to the set A, while $\mu_A(x) = 1$ denotes that element x belongs to the set A. Universe X is almost always the set of real numbers.

Definition 2 The set $x \in X \mid \mu_A > 0$ is called the support of A (*supp(A)*) and the set $\{x \in X \mid \mu_A = 1\}$ is called its kernel (*ker(A)*) [5]

B. Fuzzy ordering calculation

In 2008, fuzzy orderings were proposed by Bodenhofer in [6]. Here we recall some basic definitions used in our research; for a more extensive description see [6].

Definition 3 Consider a fuzzy equivalence relation, *T*-equivalence $E: X^2 \to [0,1]$ and a direct fuzzification, *T-E*-ordering $L: X^2 \to [0,1]$. Then, for given fuzzy set $A \in \mathcal{F}(X)$, where $\mathcal{F}(X)$ is a fuzzy superset of *X*. The fuzzy sets 'at least *A*' and 'at most *A*' (with respect to *L*), abbreviated ATL(*A*) and ATM(*A*), respectively, are defined as follow (for all $x \in X$):

$$ATL(A)(x) = \{T(A(y), L(y,x)) \mid y \in X\} \quad (1)$$

$$ATM(A)(x) = \{T(A(y), L(x,y)) \mid y \in X\} \quad (2)$$

ATL(*A*) is the smallest fuzzy superset of *A* that has a non-decreasing membership function with respect to *L*, while ATM(*A*) is the smallest fuzzy superset of *A* that has a non-increasing membership function with respect to *L*.

When *L* is a crisp ordering, the notations LTR(*A*) and RTL(*A*) are used instead of ATL(*A*) and ATM(*A*), respectively. LTR(*A*) stands for left-to-right closure and RTL(*A*) stands for right-to-left closure. The operator $\preccurlyeq$ is referred to crisp ordering.

$$LTR(A)(x) = \{A(y) \mid y \in X \wedge y \preccurlyeq x\} \quad (3)$$

$$RTL(A)(x) = \{A(y) \mid y \in X \wedge x \preccurlyeq y\} \quad (4)$$

First we describe a well-known ordering procedure for real intervals.

$$[a,b] \leq_I [c,d] \Leftrightarrow a \leq c \wedge b \leq \mathrm{d} \quad (5)$$

Equation (5) states that the only case that yields "true" or 1 value is $a \leq c$ and $b \leq d$. The inequality $a \leq c$ means that there are no elements of set [c, d] that are below the entire interval [a, b] and the inequality $b \leq d$ means that there are no elements of [a, b] that are completely above [c, d]. Equation (5) can be generalized to arbitrary crisp subsets of an ordered set $(x, \preccurlyeq)$ as follow:

$$M \preccurlyeq_I N \Leftrightarrow \big((\forall x \in N)(\exists y \in M) y \preccurlyeq x\big) \wedge \big((\forall x \in M)(\exists y \in N) x \preccurlyeq y\big) \quad (6)$$

By using the operators LTR and RTL, and considering a crisp ordering $\preccurlyeq$ on *X*, the following equivalences that hold for all $M, N \subseteq X$ are proved.

$$LTR(M) \supseteq LTR(N) \Leftrightarrow (\forall x \in N)(\exists y \in N)\, y \preccurlyeq x \quad (7)$$

$$RLT(M) \subseteq RTL(N) \Leftrightarrow (\forall x \in M)(\exists y \in N)\, x \preccurlyeq y \quad (8)$$

Since the operators LTR and RTL can be applied for fuzzy sets, an ordering of fuzzy sets $A, B \in \mathcal{F}(X)$ with respect to crisp ordering $\preccurlyeq$ is generalized as:

$$A \preccurlyeq_I B \Leftrightarrow (LTR(A) \supseteq LTR(B) \wedge RTL(A) \subseteq RTL(B)) \quad (9)$$

The inclusion $LTR(A) \supseteq LTR(B)$ means that the left flank of *A* is to the left of the left flank of *B* while $RTL(A) \subseteq RTL(B)$ means that the right flank of *A* is to the left of the right flank of *B*.

Considering fuzzy orderings above, the fuzzy ordering calculation can be determined by considering horizontal positions of comparing fuzzy sets. If the assertion (9) is fulfilled in both conditions, the fuzzy ordering value is *true* or 1. Otherwise, the operation returns *false* or 0. Figure 2 shows the comparison of fuzzy sets that yields value 1.

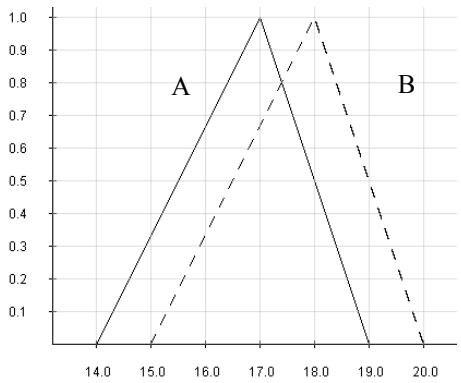


Figure 2. Comparison of fuzzy sets that satisfy (2)

From assertion (9) can be concluded that if only one condition is satisfied, it means that fuzzy sets cannot be compared - incomparable case. In this case, the fuzzy ordering operation will return *incomparable* or 0.5. Figure 3 shows the incomparable fuzzy sets.

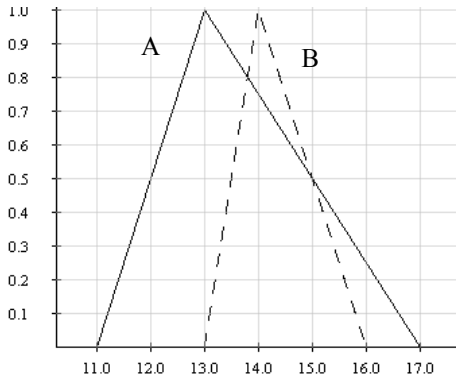


Figure 3. Incomparable fuzzy sets

Another incomparable case is the comparison of fuzzy sets having different heights. However, Skrbic and Rackovic proposed an idea to eliminate this problem in [5]. Fuzzy set A' is introduced as:

$$\mu_{A'} = \begin{cases} 1, \mu_A(x) = h(A) \\ \mu_A(x), otherwise \end{cases} \quad (10)$$

In this way, fuzzy relational operator $\leq_F$ is introduced by:

$$A \leq_F B \Leftrightarrow A' \leq'_F B' \quad (11)$$

Definition 4 Let A and B be two fuzzy sets over universe X. Order $\leq'_F$ over the set of all fuzzy sets over universe X, $\mathcal{F}(X)$ is defined by:

$$A \leq'_F B \Leftrightarrow (LTR(B) \subseteq LTR(A) \wedge RTL(A) \subseteq RTL(B)) \quad (12)$$

In the same way as with operators $<$ and $>$ on crisp domain, other relational operators, like $<_F$ and $>_F$ can be derived using the $\leq_F$ order.

C. Algorithm

As mentioned before, we consider five types of fuzzy set. Each type has different attributes that depict its properties. For example, a triangle fuzzy number contains three attributes (*LeftOffset*, *Maximum* and *RightOffset*). The *LeftOffset* refers to the beginning location of the support (*supp* in Definition 2) of fuzzy set (*LeftOffset*, 0). The *Maximum* refers to a location of its kernel (*Maximum*, 1) and the *RightOffset* refers to the end location of the support of fuzzy set (*RightOffset*, 0). Table I shows attributes for each type of characteristic function.

Attributes of fuzzy sets are used to calculate fuzzy relational operator values. Comparing two fuzzy sets, A and B, focuses on beginning, maximum and ending locations of A and B. For example, in Figure 1, two triangle fuzzy sets, A and B, are compared by operator $<$, the algorithm starts from comparing the *Maximum* attributes. If $Maximum_A$ is greater than $Maximum_B$, the result is 0 and the process ends. If not, the *LeftOffset* attributes will be compared. If $LeftOffset_A$ is not greater than $LeftOffset_B$, the process is still going onto compare *RightOffset*. If $RightOffset_A$ is greater than $RightOffset_B$, the result value is 0.5 (incomparable). If not, the result value is 1 (true). If $LeftOffset_A$ is greater than $LeftOffset_B$, $RightOffset_A$ and $RightOffset_B$ are compared. If $RightOffset_A$ is greater than $RightOffset_B$ then the result value is 0 (false), otherwise, the result value is 0.5 (incomparable). The algorithm for comparing two triangle fuzzy sets is shown in Listing 1.

TABLE I.
ATTRIBUTES OF EACH CHARACTERISTIC FUNCTION

Characteristic function	Attributes (A, μ_A)	Abbreviation
Triangle fuzzy number	LeftOffset $(A, 0)$	T-LO
	Maximum $(A, 1)$	T-MX
	RightOffset $(A, 0)$	T-RO
Trapezoidal fuzzy number	LeftOffset $(A, 0)$	TR-LO
	LeftMaximum $(A, 1)$	TR-LMX
	RightMaximum $(A, 1)$	TR-RMX
	RightOffset $(A, 0)$	TR-RO
Right shoulder	ZeroPoint $(A, 0)$	S-ZP
	Maximum $(\infty, 1)$	S-MX
Left shoulder	Maximum $(0, 1)$	S-MX
	ZeroPoint $(A, 0)$	S-ZP
Interval	LeftMaximum $(A, 1)$	I-LMX
	RightMaximum $(A, 1)$	I-RMX
Crip value	X (A)	C-X
	Y (μ_A)	C-Y

D. Crisp value

Unlike other fuzzy sets, the crisp value is a paired-value (A, μ_A). A comparison between crisp value and other fuzzy sets needs a special method.

For a relational operation between crisp value and another fuzzy set, we compare the value of attribute X of crisp value and boundary values of the compared fuzzy set. If a value X is less than the lower bound of the compared fuzzy set, the fuzzy ordering value is 1. If a value X is inside the boundary, the result value is 0.5. Otherwise, the result value is 0.

Comparing between crisp values is done in the same manner. For ordering between crisp values, A and B, following applies, if value X_A is not greater than value X_B, the result is 1. Otherwise the result is 0.

Listing 1. Algorithm for calculating fuzzy ordering between a triangle fuzzy number and another triangle fuzzy number.

```
Algorithm IsLessThan (FuzzyTriangle A, FuzzyTriangle B)
01. Compare Maximum_A and Maximum_B
02. If Maximum_A greater than Maximum_B
03.     Result is 0
04. Else
05.     Compare LeftOffset_A and LeftOffset_B
06.     If LeftOffset_A not greater than LeftOffset_B
07.         Compare RightOffset_A and RightOffset_B
08.         If RightOffset_A greater than RightOffset_B
09.             Result is 0.5
10.         Else
11.             Result is 1
12.         End if
13.     Else
18.         Compare RightOffset_A and RightOffset_B
19.         If RightOffset_A greater than RightOffset_B
20.             Result is 0
21.         Else
22.             Result is 0.5
23.         End if
23.     End if
24. End if
```

III. IMPLEMENTATION

To support our ideas, we developed the application that has two functions: manual fuzzy ordering testing and random fuzzy ordering testing. The manual testing function is used for a single test. In this case, the user can specify types of fuzzy sets and their attributes. When the process is done, the application shows an image of specified fuzzy sets and their fuzzy ordering value. Figure 4 illustrates the user interface for the manual testing function. The random testing function randomly generates comparison cases. In this function, the user can indicate types of fuzzy sets, number of generated cases, and boundary values. Figure 5 shows the user interface of the random testing function.

This application was developed on Java platform with the use of PostgreSQL to store fuzzy set attributes and cases of the random testing function.

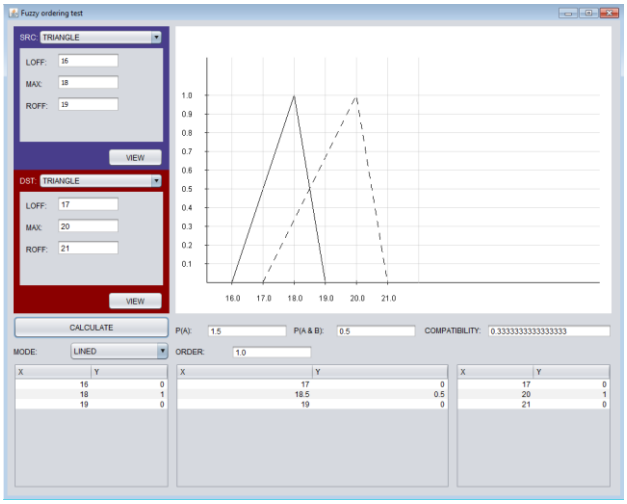


Figure 4. User interface of manual testing function

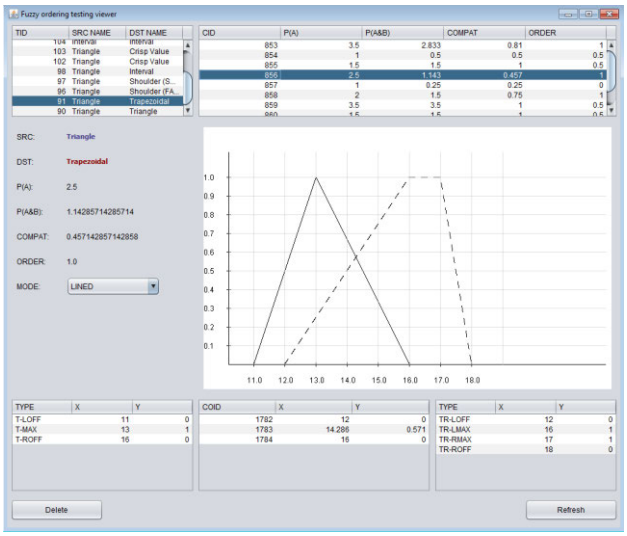
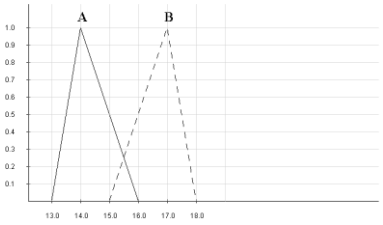
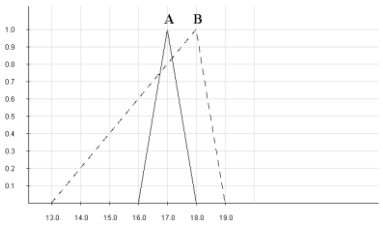
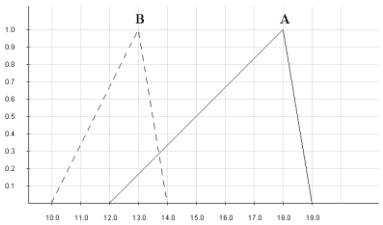
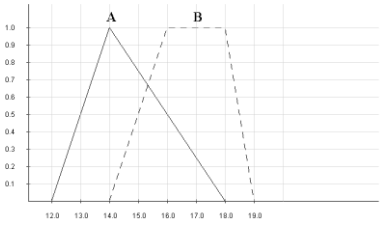


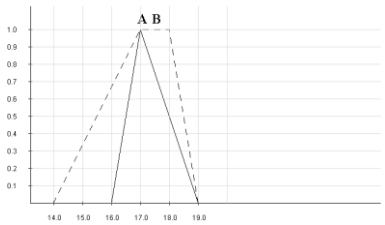
Figure 5. User interface of automated random testing function

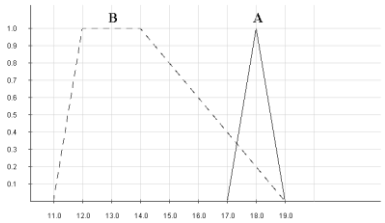
IV. TESTING RESULTS

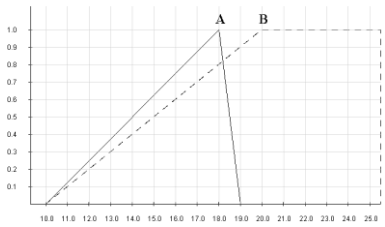
To prove the reliability of our proposed algorithms, some fuzzy ordering cases are generated randomly using random fuzzy ordering testing function of our application. As mentioned above, this paper considers five types of fuzzy sets. To cover all types of comparisons, each characteristic function is compared with the other four types including itself. Since there are two types of fuzzy shoulder, there are 36 comparison pairs. For a better variety in the comparison, the number of generated cases is set to be 10. Consequently, each pair has 10 cases of fuzzy ordering testing. Totally, there are 360 fuzzy ordering cases in our experimental results. The generated fuzzy sets are forced to position inside a boundary that is specified by the user. If the boundary is too wide, the fuzzy sets can be positioned too far from each other and have no incomparable cases. To avoid this problem, the lower bound and the upper bound are set to be 10 and 20, respectively. For the sake of brevity, some selected comparison cases between fuzzy triangle and other types are represented in Table II.

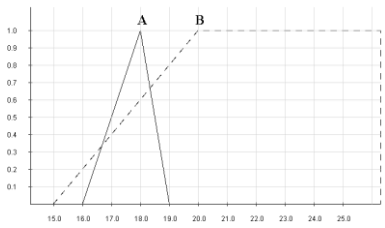
TABLE II.
A COMPARISONS BETWEEN FUZZY TRIANGLE AND OTHER TYPES

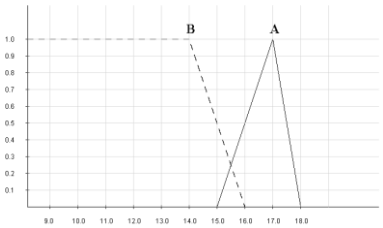
Case No.	Type of fuzzy set A		Type of fuzzy set B		Result
	Attributes	Value	Attributes	Value	
1	Fuzzy triangle		Fuzzy triangle		1
	$T\text{-}LO_A$	13	$T\text{-}RO_B$	15	
	$T\text{-}MX_A$	14	$T\text{-}MX_B$	17	
	$T\text{-}RO_A$	16	$T\text{-}RO_B$	18	
					
2	Fuzzy triangle		Fuzzy triangle		0.5
	$T\text{-}LO_A$	16	$T\text{-}RO_B$	13	
	$T\text{-}MX_A$	17	$T\text{-}MX_B$	18	
	$T\text{-}RO_A$	18	$T\text{-}RO_B$	19	
					
3	Fuzzy triangle		Fuzzy triangle		0
	$T\text{-}LO_A$	12	$T\text{-}RO_B$	10	
	$T\text{-}MX_A$	18	$T\text{-}MX_B$	13	
	$T\text{-}RO_A$	19	$T\text{-}RO_B$	14	
					
4	Fuzzy triangle		Fuzzy trapezoidal		1
	$T\text{-}LO_A$	12	$TR\text{-}LO_B$	14	
	$T\text{-}MX_A$	14	$TR\text{-}LMX_B$	16	
	$T\text{-}RO_A$	18	$TR\text{-}RMX_B$	18	
			$TR\text{-}RO_B$	19	
					

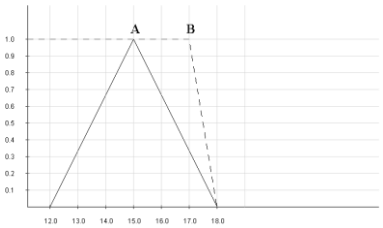
5	Fuzzy triangle		Fuzzy trapezoidal		0.5
	$T\text{-}LO_A$	16	$TR\text{-}LO_B$	14	
	$T\text{-}MX_A$	17	$TR\text{-}LMX_B$	17	
	$T\text{-}RO_A$	19	$TR\text{-}RMX_B$	18	
			$TR\text{-}RO_B$	19	
					

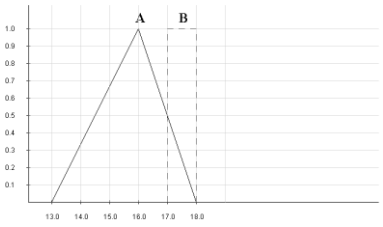
6	Fuzzy triangle		Fuzzy trapezoidal		0
	$T\text{-}LO_A$	17	$TR\text{-}LO_B$	11	
	$T\text{-}MX_A$	18	$TR\text{-}LMX_B$	12	
	$T\text{-}RO_A$	19	$TR\text{-}RMX_B$	14	
			$TR\text{-}RO_B$	19	
					

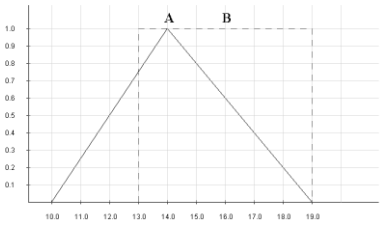
7	Fuzzy triangle		Right shoulder		1
	$T\text{-}LO_A$	10	$S\text{-}ZP_B$	10	
	$T\text{-}MX_A$	18	$S\text{-}MX_B$	20	
	$T\text{-}RO_A$	19			
					

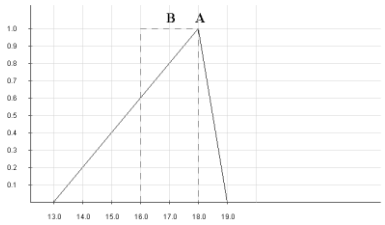
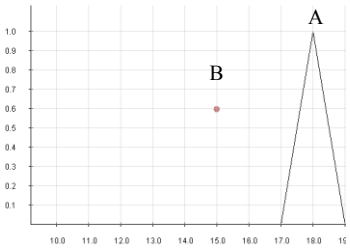
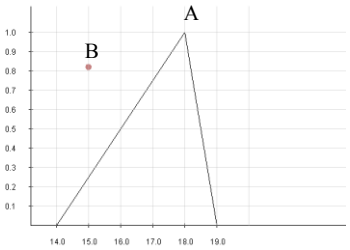
8	Fuzzy triangle		Right shoulder		0.5
	$T\text{-}LO_A$	16	$S\text{-}ZP_B$	15	
	$T\text{-}MX_A$	18	$S\text{-}MX_B$	20	
	$T\text{-}RO_A$	19			
					

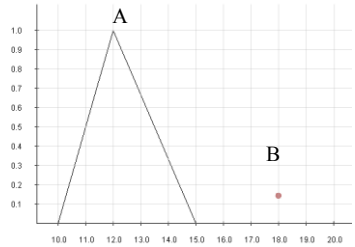
9	Fuzzy triangle		Left shoulder		0
	$T\text{-}LO_A$	15	$S\text{-}MX_B$	14	
	$T\text{-}MX_A$	17	$S\text{-}ZP_B$	16	
	$T\text{-}RO_A$	18			
					

10	Fuzzy triangle		Left shoulder		0.5
	$T\text{-}LO_A$	12	$S\text{-}MX_B$	17	
	$T\text{-}MX_A$	15	$S\text{-}ZP_B$	18	
	$T\text{-}RO_A$	18			
					

11	Fuzzy triangle		Interval		1
	$T\text{-}LO_A$	13	$I\text{-}LMX_B$	17	
	$T\text{-}MX_A$	16	$I\text{-}RMX_B$	18	
	$T\text{-}RO_A$	18			
					

12	Fuzzy triangle		Interval		0.5
	$T\text{-}LO_A$	10	$I\text{-}LMX_B$	13	
	$T\text{-}MX_A$	14	$I\text{-}RMX_B$	19	
	$T\text{-}RO_A$	19			
					

13	Fuzzy triangle		Interval		0
	$T\text{-}LO_A$	13	$I\text{-}LMX_B$	16	
	$T\text{-}MX_A$	18	$I\text{-}RMX_B$	18	
	$T\text{-}RO_A$	19			
					
14	Fuzzy triangle		Crisp value		0
	$T\text{-}LO_A$	17	C-X	15	
	$T\text{-}MX_A$	18	C-Y	0.595	
	$T\text{-}RO_A$	19			
					
15	Fuzzy triangle		Crisp value		0.5
	$T\text{-}LO_A$	14	C-X	15	
	$T\text{-}MX_A$	18	C-Y	0.819	
	$T\text{-}RO_A$	19			
					

16	Fuzzy triangle		Crisp value		1
	$T\text{-}LO_A$	10	C-X	18	
	$T\text{-}MX_A$	12	C-Y	0.143	
	$T\text{-}RO_A$	15			
					

V. CONCLUSION

This paper proposes the algorithm for binary fuzzy relational operators, which can be used to compare two fuzzy sets. Algorithms used to calculate fuzzy relational operator values are introduced. We developed an application that provides GUI and fuzzy relational operator calculations to prove the reliability of our algorithms. The testing results are generated randomly by this application. The results show that various comparisons are proved to be calculated correctly by our implementation. The proposed algorithms for fuzzy ordering will be used in FXI to enable comparison of two fuzzy sets.

Future research in this direction will tackle problems related to the implementation of aggregate functions, like MIN MAX, and SUM, using the proposed algorithms in FXI.

REFERENCES

[1] A. Champi, E. Damiani, S. Guinea, S. Marrara, G. Pasi, and P. Spoletini, "A Fuzzy Extension for the XPath Query Language" in *Flexible Query Answering Systems, Lecture Notes in Computer Science*, vol. 4027, 2006, pp. 210—221.

[2] E.J.T. Fredrick, and G. Radhamani, "Fuzzy Logic Based XQuery operations for Native XML Database Systems," in *International Journal of Database Theory and Application*, vol. 2, pp. 14—20.

[3] S. Skrbic, M. Rackovic, and M. Takaci, "Prioritized Fuzzy Logic Based Information Processing in Relational Databases," in *Knowledge-Based Systems*, vol. 38, 2013, pp. 62—73.

[4] P.S. Ueng, and S. Skrbic, "Implementing XQuery Fuzzy Extensions Using a Native XML Database," in *Proceeding of 13th IEEE International Symposium on Computational Intelligence and Informatics*, 2012, pp.305—309.

[5] S. Skrbic, and M. Rackovic, *Fuzzy databases*, Faculty of Sciences, University of Novi Sad, Novi Sad, 2013.

[6] U. Bodenhofer, "Orderings of Fuzzy Sets Based on Fuzzy Orderings Part I: The Basic Approach," in Mathware & Soft Computing, 2008, pp.201—218.

A performance analysis of the R language and an assessment of the capabilities for its improvement

Lidija Fodor*, Srđan Škrbić*
* University of Novi Sad/Faculty of Science/Department of Mathematics and Informatics, Novi Sad, Serbia
lidija.fodor@dmi.uns.ac.rs, srdjan.skrbic@dmi.uns.ac.rs

Abstract— **R is considered both as a programming language and an environment for statistical computing. It provides a wide variety of functionalities and almost limitless opportunities to extend. Its easiness to use, in combination with the interpreted nature, ensures its popularity in different areas of business and science. However, when it comes to more demanding computations in terms of performance, R turns to be significantly slower, compared to other languages. As the need for data analysis increases rapidly, the existence of a simple and at the same time efficient tool is of inestimable importance. This paper strives to present an analysis of performance drawbacks of R, and exposes an idea for possible improvements.**

I. INTRODUCTION

R is an interpreted language, used for data analysis in various fields. The language was designed in 1993 as a successor to S, by Ross Ihaka and Robert Gentleman [7]. In 1995, R was released under a GNU license. The language covers a wide range of functionalities, grouped into packages. CRAN [17] and Bioconductor [18] are well-known, commonly used repositories of R packages.

The use of R spreads through different areas of science including bioinformatics, mathematics, data-mining. On the other hand, it also finds its application among others in finance, telecommunications, pharmaceuticals, commonly used as a tool for data-mining, graphs construction and forecasting. Many of the world-wide known companies base their business decisions on the results of analysis, conducted by R.

The performance of R is partly influenced by the constructions used in the program, but they are also the results of some design decisions behind the language. We will discuss the best and most efficient ways of writing programs in R, but the focus will be on performance issues of the language itself.

In this paper, we will make the following contributions:

- Performing an analysis of R's performance, through concrete, real life examples, with comparison to other languages.
- Indicating how these problems could be solved, using modern concepts of building new language implementations.

Our goal is to compare the capabilities of R to other languages and tools as Java and MATLAB, based on practically usable examples. This approach reflects the main advantages of R, related to its conciseness and easiness to use. These are the features that established the place of R in the world of data analysis. We will show that in some cases, the performance of R can be much better, with regard to mentioned compared languages. Still, when it comes to large amount of data being analyzed or the use of complicated algorithms that require nested loops semantics, the performance issues become evident.

The reasons for inefficient program execution can be relatively easily defined ,and we will try to identify them in the fourth section of this paper. However, the solution that offers the same advantages as R, without significant syntax changes, that ensures fast programs, is still an open question that many developers are interested in.

II. RELATED WORK

A tendency to increase the execution speed of R programs arises naturally. Hence, several authors proposed some innovative ideas, about performance improvement in R.

Morandat et al. made a detailed evaluation of the design of the language [1]. They made a formal report of the semantics of the core of R and also introduced the TraceR framework for analysis, with an implementation and language evaluation review.

An interesting initiative is the work of Eddelbuettel and Sanderson [3]. They use the C ++ language with extensions in form of packages, called Rcpp, relying on Armadillo C ++ library, in order to work with matrices. In addition, they apply a template-based framework for metaprogramming, with the aim to easily convert R's linear algebra algorithms to C++.

Li et al. introduce the principle of automatic and transparent parallelization, using a runtime framework called pR [5]. The idea of semi-automatic parallelization also occurs in the work of Jiang et al. [6]. This approach introduces the idea of adding pragma directives to the source code in OpenMP style.

An equally interesting approach is program specialization at the level of the interpreter, developed by Wang et al. [4]. The idea relies on a direct optimization of the virtual machine using R's extensions ORBIT VM, in order to avoid allocation, by the principle of aggressive deletion of allocated objects and shortening the instructions paths.

FastR is a project, developed by Kalibera et al., which represents a unique subset of the implementation of the R language, built within the host language Java [2]. This approach uses the principle of an AST interpreter, and is based on ANTLR parser generator, which creates a tree, on the basis of the source code. The execution is achieved by an executable tree, while the conversion between the trees happens during an evaluation process.

We propose an idea for solving the performance issues of the language by developing a completely new

implementation, without changing its syntax. However, the underlying implementation should provide optimizations and possibly parallelization, in such a way that the user should not be aware of them.

III. KEY FEATURES OF R

R is a functional, object-oriented and dynamic language. This unique combination of features makes the language flexible and plain to the end users. Its interpreted nature, combined with the wide range of existing functions, makes it possible to perform various complex computations with a single function call.

A. Functional nature

In accordance with its functional features, R treats functions as priority entities, which makes it possible to assign them to variables or pass them as arguments to another functions. From the aspect of scopes, R uses lexical scoping, so the range of each value can be determined statically, before run-time. The global scope, called "environment", can contain implicitly created sub scopes for the need of loops and functions, but explicit user-defined sub scopes are also allowed. In addition, dynamic binding allows the insertion of names to scopes dynamically.

Function arguments are evaluated lazily, by packing them into so called "promises", containing the argument's expression and the information about the current scope. That way, promises are evaluated as they become needed.

R provides high-level flexibility, when declaring function arguments. First of all, default values are enabled. Besides that, a variable number of parameters can be also specified, treated as an arbitrary size array. When calling functions, the parameters can be passed positionally, named or combined, with the restrictions that each call should contain at least one positional parameter, as well as that after a variable list of arguments, the named approach should be used.

These items provide a convenient tool for the end users to work with functions, without deep understanding of how they actually work.

B. Dynamic nature

Dynamic language features are manifested through dynamic evaluation and dynamic type system. The dynamic type system frees the source code from explicit type declarations. As a consequence, the same variable can hold values of different types at different places in the code. Assigning a type to a value, as well as conversions between them are performed implicitly. R supports numeric, logical and character as primitive types.

Dynamic evaluation is supported through the function "eval", and the reverse process of converting an evaluated expression to its string form is also present. As mentioned before, R enables reflection over environment, by means of direct access and modification from the program. This creates an excellent base for flexible program manipulation and extensions creation.

C. Object-oriented nature

R supports two different object models, called S3 and S4. S3 is the older approach and is also called "single-dispatch generic function system". It is not an object-oriented system in the true sense, as it does not introduce classic object-oriented concepts. Instead, it relies on setting class attributes for variables. It should be mentioned, that variables can have different attribute values as their integral parts, which describe some features of the variable, like size or type. The class attribute only serves to define methods over the class. When a method is called for a variable with a defined class attribute, the method defined for that class will be executed.

The S4 object model is also called "classes and multi-methods system". This is a newer approach, with all the standard object-oriented features. It allows declaring classes with "slots" for holding values, methods, as well as defining hierarchies of inheritance.

Although the S3 system has greater popularity due to its simple nature and longer history, the S4 system brings the real object-oriented note to the language, and opens further possibilities for program creation.

D. Basic elements of R programs

The basic data type in R is the vector type. Each primitive value is treated as a single-element vector. Simple binary operators are applicable to vectors, with the semantics of applying the operator element-wise. The length of one vector should be a multiple of the length of the another, but the operator can be applied even if this condition is not satisfied. The main feature of vector type is that all its elements have to be of the same type. R also supports the value NA (Not Available), which is important in statistical computations. Matrices, lists, data-frames, environment and functions are considered also as non-primitive data types.

While matrices have the usual meaning, lists should be explained more carefully. In R, they represent heterogeneous vectors, which mean that they can hold different number of vectors of different sizes and different types. This makes them extremely flexible for representing heterogeneous data. Data-frames are similar to lists, but restricted to vectors of the same size.

R supports common control structures, as the if branching and iterations in forms of for, while and repeat loops. As an alternative to loops, R introduces a family of "apply" functions. These functions can be used for different data structures, by applying some arbitrary function on the structure. Apply functions are the preferred way for iterating over data-structures, as they result with better performance than the direct use of loops.

To perform a piece of job, R commands (assignments and calls), can be grouped inside scripts, or can be directly entered to R's command line one by one. Extending the present possibilities of the language can be easily achieved by creating the desired functions and putting them to packages.

The R community mainly includes end users that interactively calls existing functions for reading a piece of data, performing some analysis and graphically represent the results. From this aspect, the clearness and simplicity of the language are the most important demands.

There are two more groups of R users: experts from the field of statistics and programmers. The wide range of statistical algorithms is available thanks to statisticians. On the other hand, programmers are the group with deeper understanding of the features of the language.

The question is whether it is sufficient to write as much as possible optimal functions at background, so that they could perform efficiently when the users at the front end use them. The answer is negative. A large number of R's functions is written with care, to the extent that they are implemented in C programming language, in order to gain as much as possible speedup. Despite that, they do not show significant execution time improvements. The reasons for this and some other performance problems will be described below.

IV. PERFORMANCE

A. Basic notes

Despite the described advantages of R, working with large data sets shows serious deficiency in performance. Earlier studies emphasized some of the weakest points of the language. During their detailed language design analysis, Morandat et al. evaluated performance related problems [1].

Firstly, a significant amount of time is spent on memory management. The reasons for this are linked to the basic features of the language: allocating space for vectors, copying arguments, garbage collection, built-in functions calls. Copying arguments is the result of passing by value, so regardless of the size of the structure that represents the argument, a copy is created. Also, all the functions that work with values defined in an upper environment, have their own local copies of the values. A certain amount of time is spent on lookup and pairing values with arguments, hence the language is interpreted with dynamic binding.

An interesting fact is that a large number of provided R functions, is written in another languages, usually C and Fortran, due to performance issues. While these functions contribute to a more efficient execution, the amount of time spent calling them is greater than it might be expected. On average, a fifth of execution time is spent on calling such functions.

Memory consumption is another important aspect that influences efficiency. Beside the process of allocation for user defined data, a significant amount of memory is allocated internally for the interpreter. Allocating a vector for a simple value leads to unnecessary occupation of memory and to time consumption for deallocation later. Allocation and deallocation happen on the heap, which further affects performance. These are small components of the total execution time and may seem insignificant, but their combination and multiple application can seriously impact performance.

The dynamic nature of the language directly reduces the possibilities for optimizations. The advantages in terms of simplicity and clarity certainly have a cost in terms of speed. The combination of a loop and dynamic type system can lead to inefficient running of the program.

```
x<-5
a<-c()
for(i in 1:10000){
  a[i]<-i*x
}
print(a)
```

Listing 1 – Illustration of the dynamic typing problem.

Consider the small example, given in Listing 1. As the variable x is not declared with type, each iteration of the for loop has to decide on which data type should it apply the multiplication operator. This is a trivial case, but thinking of more complicated situations, which require working with methods in object-oriented concepts, illustrates the problem even more.

While declaring functions, one should carefully decide what is the minimal number of arguments needed, as more arguments can lead to slower execution. The reason for this is the lazy evaluation, which packs the arguments into promises. Practically, the moment of the evaluation of promises often happens immediately, so the costs of creating the promises affect the program. Even with the simplest functions that just apply a binary operation, each additional argument increases the execution time by a few milliseconds.

The costs of calling functions can be also an issue. The interpreter needs to create a scope for the function call, copy the values and add the arguments to the scope. The fact that arguments can be named, and that variable parameter lists could be used, requires additional effort.

B. The preffered ways of writing programs

Although the nature of R is the main reason for performance issues, it is worth to mention that the right way of writing programs inside it can decrease the execution time to certain degree.

When operating on vectors, it is important to keep in mind that loops perform inefficiently. Vectorized functions can, on the other hand, operate on vector elements, with the efficiency, as with primitive values. Let's consider the example of creating a vector, containing prime numbers in range of 1 and 100000.

Listing 2 shows the source code using a for loop, while Listing 3 illustrates the use of vectorized function. On a Quad core AMD Phenom 9550 machine, with 2.2 GHz and GNU/Linux(x86-64), used for all examples here, the execution time of the first version of the program is 0.8 seconds. Running the second approach gives an execution time of 0.12 seconds. It can be noticed that the vectorized approach is nearly 6 times faster. For more complicated uses, the difference could be even larger.

Another important note is that recursion should be avoided, hence R does not support tail recursive calls optimizations. Also, anonymous functions should be used instead of named ones, where possible.

```
library(gmp)
v<-1:100000
w<-c()
for(i in seq(along=v)){
  if(isprime(v[i]))
    w[length(w)+1]<-v[i]
}
```

Listing 2 – Prime numbers, loop-based approach.

```
library(gmp)
v<-1:100000
w<-v[isprime(v)!=0]
```

Listing 3 – Prime numbers, vectorized approach.

Respecting these rules, the real performance of R programs can be observed, as deficiency caused by bad style of programming is eliminated.

C. Case-studies

As R is often used for data-mining, consider its application when working with time-series data. Listing 4 shows the use of hierarchical clustering on time-series data, loaded from a file, applying the similarity measure DTW and plotting the resulting dendrogram. The library DTW contains the function for applying dynamic time warping (dtw) similarity measure. The test file contains 100 time series, each 5000 points long. The values of points are between 0 and 1.

Reading the data is achieved by simply calling the function read.table, while the distance matrix is created by calling dist. The results are recorded in a csv file. The function hclust is used to perform hierarchical clustering, and the results are used to create a dendrogram. The code is self-descriptive and very easy to understand. The whole process of reading the data, performing clustering, and writing the results into a file is achieved by several function calls.

It is clear, that the program is written respecting the previously described recommendations for writing the most efficient code. The code is mainly consisting of predefined function calls. The execution time of this program is 18.19 seconds. This is quite a large number for a dimension of 5000 points of time-series. Practically, it is often necessary to work with much larger data series. To identify time consumption rates of different parts of the program, we will try to remove particular calls and compare the timings. If the dendrogram creation is eliminated, the execution time decreases to 6.26 seconds, which indicates how costly the plotting can be.

```
start<-proc.time()
library(dtw)
name<-"TestData.csv"
fName<-substr(name,0,nchar(name)-4)
data<-read.table(name, header=F, sep=";",
                 row.names=1)
distMatrix<-dist(data, metod="DTW")
write.table(as.matrix(distMatrix),
            paste(fName, "DTW", ".csv",
                  sep=""))
hc<-hclust(distMatrix, method="average")
jpeg(paste(fName,"DTWavg",".jpg", sep=""))
plot(hc, main=paste(substr(fileName, 0,
                    nchar(fileName)-4),
                    "DTW average",
                    "linkage", sep=" "))
dev.off()
end<-proc.time()
elapsed<-end-start
print(elapsed)
```

Listing 4 – Case study, time-series data-mining.

Secondly, if the reading from a file is replaced by creating a matrix filled with random data, the time will be 0.95 seconds. The conclusion is that data input/output represents the main problem in this kind of application. Data-mining always require a lot of communication with files, so the time spent for that is usually not negligible.

However, comparing the same time-series data mining program, written in Java and MATLAB, one can make the conclusion, that R is the best choice, regarding to the amount of code needed for the implementation, and also regarding to the execution time. Figure 1 shows the amount of code needed in R, Java and MATLAB for the same job, and Figure 2 demonstrates the execution times in each of them. Despite of the performance issues, it is obvious that R is a better choice for data-analysis than a general purpose high-level language Java, requiring a lot of programming effort to achieve the same result as with 20 lines of R code. From this perspective, R turns to be also better than MATLAB, as it performs faster.

Consider one more example, shown on Listing 5. This is just an example, illustrating computations performed on data, stored inside a data-frame. It is often convenient to use tabular representation of the data.

One of the most often approaches is to use one of the columns as a measure for grouping, and then perform calculations on the rest of the data.

For example, all the data about the employees in a company can be represented by a data-frame. It may be needed to calculate the average salary or average number of work hours for each position in the company. For these, a grouping based on employee's position should be made first. Then, for each group, the average should be calculated. Of course, it is likely that the algorithms, that need to be applied to grouped data, can be much more complex. Listing 5 illustrates that situation. A data frame with 3 columns and 1000 elements is created. The third column of the data-frame contains values 1 and 2, alternately, so it is a good base for grouping the data into two groups.

The function aggregate can be applied to multiple columns of data-frame, using the list of groups, provided as the second parameter, and applying the function given as the third parameter. This call will apply the anonymous function to the first two columns and group the results, based on the values of the third column. The anonymous function creates a vector of 100 elements for each value, where the elements are the square root of the value.

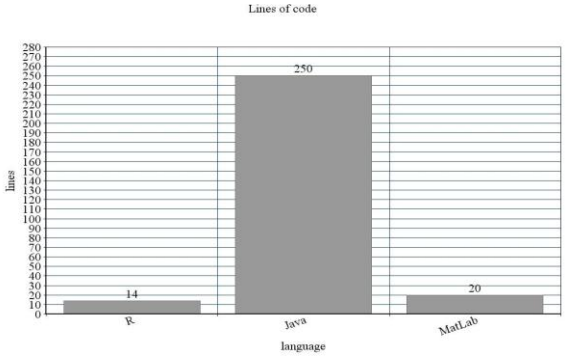


Figure 1. Execution times for time-series mining case- study in R, Java and MATLAB

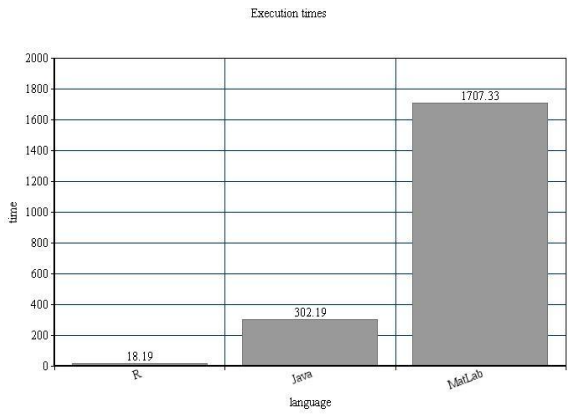


Figure 2. Lines of code needed for time-series mining case study in R, Java and MATLAB

Finally, it calculates the mean value of the numbers from the vector. The final result is a data-frame again with 2 rows and 2 columns.

The example showed the case of applying an operation on a vector for each element of data-frame. This is an implicit use of nested loops. The rules of using vectorized functions were followed, but still the execution time is 49.59 seconds. If the number of elements for vectors is decreased to 50, the timing is 12.28 seconds. So, for doubling the number of elements in vector per data-frame element, the time is 4 times longer.

Based on the examples illustrated, it can be concluded that despite of paying attention to write efficient programs, R expresses serious problems when it comes to the execution time of programs, that rely on working with large data structures, or that require complex calculations in terms of nested loops. It is clear that these concepts impair performance in other languages too, but usually not as much as here.

```
start<-proc.time()
prop1<-1:1000
prop2<-seq(2,2000,2)
prop3<-rep(c(1,2),500)
myData<-data.frame(prop1,prop2,prop3)
res<-aggregate(myData[,1:2],
list(myData$prop3),
               function(x){
                 sapply(rep(sqrt(x),100), mean)
               })
end<-proc.time()
total<-end-start
print(total)
```

Listing 5 – Case study, data-frame data analysis.

Since these cases are very common in practice, the conclusion is that the performance of R is not at an appropriate level, and that there exists a need to find an alternative solution for higher level of efficiency.

V. THE POSSIBILITIES FOR IMPROVEMENT

The need for higher performance R is evident. However, the improvement needs to retain the basic form of the language. Introducing changes to the syntax and semantics or including additional elements in the source code could lead to the loss of simplicity and clarity. Therefore, our idea is to construct a new, more efficient implementation of the language, rather than attempting to improve the existing one.

Building a language from the ground is a demanding undertaking that requires knowledge from different fields, in order to create a language specification, a compiler construction, a set of necessary libraries, an integration of optimizations, and many other aspects.

This complexity can be overcome, by means of modern approaches to build embedded languages within a host language. To achieve this, it is very important to choose a host language that widely supports the implementation of other languages inside it. Domain specific languages (DSLs) are those that are embedded inside other languages. R could be considered as a DSL as well, since it is mainly oriented to statistical calculations, although it is officially not treated so. An appropriate host language should provide a convinient way to introduce the DSLs abstractions, based on the constructions of the host language. Basic features of the host language, as its syntax clarity, extremely influence the process of embedding a new language.

Another important aspect, when building a DSL, is to pay attention to the possibilities for optimization and potentially parallelization on lower levels, without disturbing the form of the DSL. It is also very important to find a right way to overcome the high level of abstraction, as one of the main features of DSLs. Although abstraction simplifies the use of the language, it is in contradiction with the need for efficiency. There are two standard ways to overcome this barrier. The first one is the elimination of abstractions before compiling, so the compiler is not even aware of them. This can be achieved using staging. Staging or multi-stage programming performs a separation of the process of compiling into series of successive phases, which enables code generation. The second approach is to provide domain-specific knowledge to the compiler, by adding new phases of compilation and creating intermediate representations. In recent period, there appeared some approaches that can combine these two mechanisms, through the creation of intermediate representations using staging. Thus, the principles of construction of embedded DSLs, with the inclusion of optimizations, relying on generative programming through staging will be our basic idea while attempting to build a new R implementation.

VI. CONCLUSION

According to described properties of the R language, it turns out that its popularity is justified due to its simplicity and wide range of areas of use. However, when it comes to dealing with problems related to large data sets, efficiency problems arise and disable its use in such applications. We showed some of the aspects, influencing deficiency of R through examples and comparison with other languages.

The objective for further work will be to build a more efficient implementation of the language, to make a

comparative analysis of performance and evaluate the program execution. The described concepts of building DSLs will be the leading ideas. Primarily, it is necessary to identify an appropriate host language with appropriate tools that supports optimizations or generative programming. Such an approach could provide a completely new implementation of the subset of R, with significantly better performance, that could enter into wider use.

ACKNOWLEDGMENT

Results presented in this paper are the result of collaboration between Faculty of Science in Novi Sad and Ecole Polytechnique Federale de Lausanne. The work is partially supported by Ministry of Education, Science and Technological Development of the Republic of Serbia, through project no. ON174023: Intelligent techniques and their integration into wide-spectrum decision support.

REFERENCES

[1] F. Morandat, B. Hill, L. Osvald and J. Vitek, "Evaluating the Design of the R Language", in Proceedings of European conference on Object-Oriented Programming (ECOOP), 2012.

[2] T. Kalibera, P. Maj, F. Morandat and J. Vitek, "A Fast Abstract Syntax Tree Interpreter for R", in Proceedings of the 10th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments, VEE '14, New York, USA, pp.89-102, ACM, 2014.

[3] D. Eddelbuettel and C. Sanderson, "RcppArmadillo: Accelerating R with High-Performance C++ Linear Algebra". Computational Statistics and Data Analysis, 71, 2014.

[4] H. Wang, P. Wu and D. Padua, "Optimizing R VM: Allocation Removal and Path Length Reduction via Interpreter-level Specialization", in Proc. CGO, pp.295-295, 2014.

[5] J. Li, X. Ma, S. Yoginath, G. Kora and N. F. Samatova, "Transparent Runtime Parallelization of the R scripting language". Journal of Parallel and Distributed Computing, 71(2), 157-168, 2011.

[6] L. Jiang, P. B. Patel, G. Ostrouchov and F. Jamitzky, "OpenMP-style parallelism in data-centered multi-core computing with R", in Proc. PPOPP, 2012, 335-336

[7] R. Ihaka and R. Gentleman. "R: A Language for Data Analysis and Graphics". Journal of Computational and Graphical Statistics, 5 (3):299-314, 1996.

[8] T. Wurtinger, C. Wimmer, A. Woss, L. Stadler, G. Duboscq, C. Hummer, G. Richard, D. Simon and M. Wolczko, "One VM to Rule Them All", in Proceedings of Onward!, the ACM Symposium on New Ideas in Programming and Reflections on Software, 2013.

[9] R Development Core Team, "R: A language and Environment for Statistical Computing", R Foundation for Statistical Computing, 2011.

[10] R Development Core Team, "The R Language Definition". R Foundation for Statistical Computing, http://cran.r-project.org/doc/manuals/R-lang.html

[11] J. M. Chambers, "Software for data analysis: Programming with R", Springer, 2008

[12] R. Gentleman, et. al., eds. "Bioinformatics and computational biology solutions using R and Bioconductor", Statics for Biology and Health, Springer, 2005.

[13] R. Gentleman and R. Ihaka. "Lexical Scope and Statical Computing". Journal of Computational and Graphical Statistics, 9: 491-508, 2000.

[14] D. Smith. "The R ecosystem". In The UseR Conference 20011, August 2011.

[15] J. Talbot, Z. DeVito, and P. Hanrahan. "Riposite: a trace-driven compiler and parallel VM for vector code in R". In proceedings of Parallel Architectures and Compilation Techniques (PACT), 2012.

[16] L. Tierney. "A byte code compiler for R". University of Iowa, 2015. http://homepage.stat.uiowa.edu/~luke/R/compiler/compiler.pdf

[17] R project. CRAN: The comprehensive R archive network, 2015. http://cran.r-project.org

[18] Bioconductor: Open source software for Bioinformatics, 2015. http://www.bioconductor.org

The Role of Business Process Modeling in Information System Development with Disciplined Agile Delivery Approach

Ljubica Kazi*, Miodrag Ivkovic*, Biljana Radulovic*, Madhusudan Bhatt**, Narendra Chotaliya***
* University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia
** University of Mumbai, R.D. National College, Mumbai, India
*** Gujarat Government, Department of Education, Knowledge Consortium, Ahmedabad, India
ljubicakazi@hotmail.com, misa.ivkovic@gmail.com, biljana.radulovic66@gmail.com, mmbhatt@gmail.com, narendra_chotaliya@yahoo.com

Abstract— **Agile approach to software development is formally established in 2001 with Agile Manifesto promotion. Many different methods of agile approach were used by practitioners for many years and the need for their integration and tailoring has emerged to new methodology created by IBM and promoted since 2012 as disciplined agile delivery approach. This paper aim is to describe the role of modeling in disciplined agile approach. A case study is presented with results of project of information system development within educational environment. Selection of models to be used for information system development within disciplined agile approach in this case study is oriented towards business process model, UML's use case model and data models. They are represented as a basis for agile software development and iterative delivery within the case study. CASE study is focused on the analysis of the impact of business process modeling to information systems software project performance within Disciplined Agile Delivery approach.**

I. INTRODUCTION

Evolution of software engineering [1] established paradigms and appropriate methods and tools, starting with 1950's (software engineering as hardware engineering), 1960's (software crafting), 1970's (formality and waterfall model), 1980's (productivity and scalability), 1990's (concurrent vs. sequential processes, open source development and usability). 2000's are characterized [1] by development of agile methods, value-based software engineering, model driven development and integration of software and system engineering and 2010's with globalization and development of system of systems.

Developers of agile methods established Agile Software Development Alliance [2] in February 2001 and signed document "Manifesto for Agile Software Development" [3] with 4 core values and 12 principles to follow in software development. This event is considered as a formal start of application of agile methods in practice. During the process of agile methods tailoring, different terminology from diversity of agile methods confused practitioners and incompleteness of particular agile methods required integration in their practical use and delivery of results [4]. Therefore, in 2012 IBM proposed [5] Disciplined Agile Delivery framework (DAD) as a process framework that is oriented to delivery of solutions by applying integration of different agile methods. "The Disciplined Agile Delivery (DAD) decision process framework is a people-first, learning-oriented hybrid agile approach to IT solution delivery. It has a risk-value delivery lifecycle, is goal-driven, is enterprise-aware, and is scalable." [4] Basic values and principles described in [5] are formulated as Disciplined Agile Manifesto (DAM) [6].

In continuation of our related research in information systems modeling ([7], [8], [9], [10], [11], [12]), with this paper we aim to investigate the role of modeling in information system development, within a context of disciplined agile delivery approach. Change in development paradigms and methodology directed to speed in solution delivery and stakeholders satisfaction minimize efforts and time spent on modeling. These changes enforce the need for research about the position of modeling in disciplined agile delivery as contemporary industry model. Results of this research could be basis for change in higher education of information systems and software engineering teaching plans.

Basic research question of this paper is: "What is the role of modeling within disciplined agile delivery in information system development?" This question could be elaborated with answering to particular, more precise questions: Is there a need for modeling within disciplined agile delivery in information system development? Since DAD methodology denote "initial modeling", which level of details is appropriate? Is there any particular type of modeling process and types of models in DAD methodology for information systems development? b) Which models are necessary (core)? c) Which models to use in aim to minimize time and efforts in creation of models? By selecting particular types of models, is the selection appropriate? b) Which criteria could prove that the selection of models is appropriate?

There are several feasible directions in choosing research methods regarding previously set research questions: analysis of related work in literature; conducting a survey in software industry, with questions regarding their attitudes regarding position of modeling in application of DAD and impressions about their experiences in this field; empirical research within a case study on application of DAD within educational environment, with students projects implementation. In

this paper, we choose empirical research as a case study on application of DAD within educational environment.

Second section of this paper represents background about disciplined agile delivery approach. Third section represents motivation for the research that is related to position of disciplined agile delivery approach in higher education. Fourth section is particularly related to research methodology for the case study empirically conducted at University of Novi Sad, Technical faculty "Mihajlo Pupin" Zrenjanin, Serbia. Fifth section represents results of case study and discussion about research questions and hypotheses. Final section represents conclusions regarding the role of modeling in information system development and possibilities of focusing on core models related to functional and data aspect.

II. BACKGROUND

The term "agility" is defined in [13], based on definitions from [14] and [15] as "an effective integration of response ability and knowledge management in order to rapidly, efficiently and accurately adapt to any unexpected (or unpredictable) change in both proactive and reactive business / customer needs and opportunities without compromising with the cost or the quality of the product / process". Agility is closely related to flexibility and leanness, but they should be distinguished (i.e. they are concepts within the agility as a broader term). Research [13] systematized 28 frameworks and models "describing the concepts that determine agility or at least proposed different items to measure agility." [13] These frameworks could be categorized in four domains: Agile Manufacturing, Agile Software Development, Agile Organization/Agile Enterprise and Agile Workforce and all of them together include 33 different concepts that are included in agility definition and frameworks. All these concepts could be categorized in 5 domains: organizational culture, technology, workforce, customer, organizational abilities.

Agile methods application in the first period of use (before establishment of Agile Manifesto) were sometimes considered risky, being interpreted as opposed to planned and predictive process models and, therefore, leading to chaos [16]. Both agile methods and plan-driven methods have home ground where they fit best [16]. Plan-driven methods characteristics: developers (plan-oriented, adequate skills); requirements (knowable early, largely stable); architecture (designed for current and foreseeable requirements); teams (larger teams and products); primary objective (high assurance). Agile methods characteristics: developers (agile, knowledgeable, collaborative); customers (dedicated, collaborative, representative, empowered); requirements (largely emergent, rapid change); architecture (designed for current requirements); teams (smaller teams and products); primary objective (rapid value). Hybrid approaches should be considered [16] in scaling between these two extremes [17].

During period 2006-2012, the Disciplined Agile Delivery (DAD) process decision framework was developed by IBM (Scott Ambler), as the result of the observations of different agile methods application worldwide [17]. Basic characteristics of disciplined agile delivery framework are [17]:

- Hybrid – using strategies from agile methods: Scrum, Extreme Programming (XP), Agile Modeling (AM), Unified Process (UP), Kanban, Outside in Development (OID), and Agile Data (AD) etc.
- Enterprise aware – "Disciplined agile teams recognize that they are part of a larger, organizational ecosystem and act accordingly", cooperating with other teams within the organization.
- Solution focused – "from just producing software to instead providing consumable solutions that provide real business value to your stakeholders within the appropriate economic, cultural, and technical constraints. Software is clearly important, but in addressing the needs of our stakeholders we will often provide new or upgraded hardware, change the business/operational processes that stakeholders follow, and even help change the organizational structure in which our stakeholders work."
- Delivery focused – orientation to continuous delivery, throughout the lifecycle,
- Goal driven – Team should adapt to change of requirements and development priorities. "DAD's goal-driven approach underlies the idea that to be effective at applying agile a team must understand the context in which they are working… different teams face different situations; therefore they will need to adopt their strategy to reflect the situation. Each team needs to identify an initial technical strategy, explore their initial scope, develop an initial plan, and fulfill many other goals but they will achieve these goals in different ways. The DAD process framework provides straightforward guidance to help you to make these tailoring decisions effectively. It does this by explicitly describing the process decision that you are making and then walks you through the process of making it."

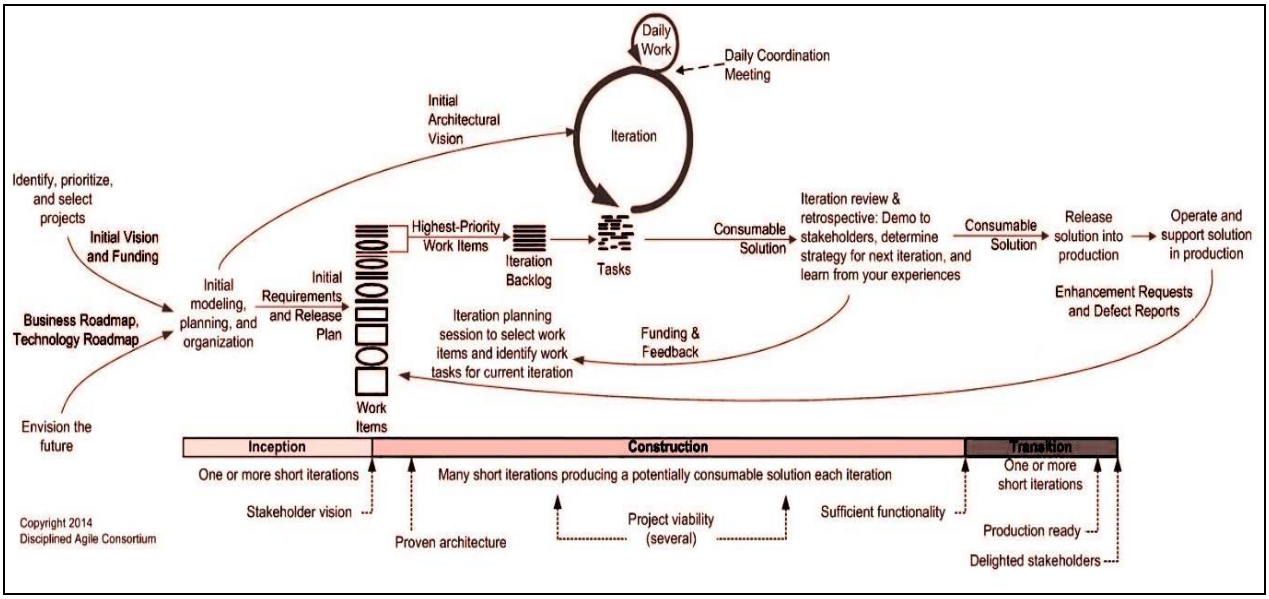


Figure 1. Disciplined Agile Delivery Life Cycle ([17] [28])

DAD life cycle is presented (in [17], Figure 1,) within steps and iterations as follows:

1. DAD lifecycle starts with identification, prioritization and selection of projects, where initial vision and funding of each project is defined.
2. Inception phase of a project goes through one or more short iterations. It consists of initial modeling, planning and organization where initial requirements and release plan is delivered and finally, stakeholder consensus is established.
3. Construction phase is done with short iterations and each iteration produces potentially consumable solution. Work items are defined and highest priority work items are selected to be included in iteration backlog. Tasks are defined upon iteration backlog. Tasks are done within iteration during each day work where each day has daily coordination meeting. After iteration is finalized, potentially consumable solution is delivered for iteration review and retrospective, demo to stakeholders is created and presented and strategy for the next iteration is determined. Iteration planning session selects work items and identifies work tasks for next iteration. Construction phase is finished when sufficient functionality is developed.
4. Transition phase starts when sufficient functionality is included in solution and solution is released in production. This phase consists of several short iterations.
5. Finally, solution is operating in working environment ("working solution in production") with support available for enhancements requests and defect reports.

III. MOTIVATION - POSITION OF DISCIPLINED AGILE DELIVERY APPROACH IN HIGHER EDUCATION

Detailed analysis of why agile software development should be included in software engineering educational programs emphasizes ten most important reasons [18]: 1. Agile was evolved and is applied in the industry; 2. Agile educates for teamwork; 3. "Agile" deals with human aspects; 4. "Agile" encourages diversity; 5. "Agile" supports learning processes; 6. Agile improves habits of mind; 7. Agile emphasizes management skills; 8. Agile enhances ethical norms; 9. Agile highlights a comprehensive image of software engineering; 10. Agile provides a single teachable framework.

In aim to align education with the needs of professional environment, agile methods are applied at higher education within undergraduate [19] and master studies [20] [21] as advanced software engineering and project management contents as well as dedicated specific courses/subjects. Recent trends in higher education of computer science introduce undergraduate capstone courses i.e. projects, which could include [22] agile methods teaching and practical work of students dealing with professional-like projects. Moreover, special tools, such as SCRUMIA [23] (a computer game) were designed in aim to enable application of agile methods and improve learning experiences. Analysis of particular courses and educational experiences of using agile methods show that different agile methods are presented theoretically, but one of the methods (such as SCRUM) is mostly practically exercised with students' teams dealing with software projects. DAD framework represents a novel approach [5] which is still not included within educational environments. Therefore, it is very important to examine elements of this approach, particularly the role of modeling, which could influence change in higher education in information systems and software engineering.

IV. CASE STUDY RESEARCH METHODOLOGY

In this paper we present empirical research in a case study that is conducted at University of Novi Sad, Technical Faculty "Mihajlo Pupin" Zrenjanin, Serbia. Basic elements of research methodology are represented as starting statements regarding previously formulated research questions and as research questions that this case study address, including methods that are used.

A. Starting statements regarding research questions

Question 1: Is there a need for modeling within disciplined agile delivery in information system development? *Statement*: There is a need for modeling in any information system development, and DAD methodology denotes that need as "initial modeling" [17]. Question 2: Which models to use in aim to minimize efforts in information systems software development? *Statement*: Basic information system development models include business process models, functional models and data models [24][25]. They are needed for any information system development.

B. Research questions in the case study

1. Since DAD methodology denote "initial modeling", which level of details is appropriate?

2. Which models are necessary (core)? Which models to use in aim to minimize efforts in creation of models?

3. By selecting particular types of models, is the selection appropriate? b) Which criteria could prove that the selection of models is appropriate?

C. Modeling process selection

Regarding previously selected research questions in this case study, particular information system software development process is selected for this case study: Requirements collection starts with collection of documentation representing the business process of an organization, functional requirements specification from clients and analysis of documentation forms. Method of structural system analysis is used for presenting business process model and data flow by using data flow diagram and data dictionary. Process tree represents a basis for mapping of primitive processes to software functions, while data dictionary elements are used as a basis for conceptual data model creation. Mapping of elements from business process models to software design is presented in [11]. After creating business process model, i.e. data flow diagram with data dictionary, UML's use case diagram is created for the functional modeling purpose, while conceptual data model, physical model and object oriented model (class diagram) is created for the data modeling purpose. All models (i.e. diagrams with additional specifications - business process model, use case model and data models) are created in CASE tool Power Designer, by using advantage of data interoperability between models (export data from business process model to conceptual data model) and advantage of automated creation of models (from conceptual data model to physical data model and object oriented model).

Obviously, business process model is core model which gives basis for functional modeling (use case model) and data modeling (conceptual and physical model, as well as object oriented model – class diagram). Main question is "*In aim to minimize time and efforts in modeling, could business process modeling be avoided? Could we just do minimal modeling in functional and data segment without previous business process modeling? What are the consequences if we omit business process modeling?"*

D. Indicators and hypothesis selection

Within the project management "iron triangle" approach, basic success factors for any project is related to scope, quality, time and resources/cost. [26] Since cost is closely related to time, and quality is closely related to scope, we select scope and time as basic quality indicators for this case study.

Research hypotheses could be formulated as:

Hypothesis 1 related to time indicator – "Business process modeling increase overall project duration".

Hypothesis 2 related to time indicator - "Omitting business process modeling increase number of development iterations in software development".

Hypothesis 3 related to scope indicator – "Business process modeling enables completeness of the project scope".

Hypothesis 4 related to scope indicator – "Omitting business process modeling gives partial results of the project scope".

E. Case Study Research Sample and Methods

In this paper, sample represent results of students' projects implemented within educational environment. There are two categories of students' projects:

1. First category are mandatory projects within practical exams at University of Novi Sad, Technical faculty "Mihajlo Pupin" in Zrenjanin. In aim to fulfill prerequisites for entering exam, students need to do their practical homework. First category of projects included business process modeling.

2. Second category projects are with optional engagement in development of information system of an educational institution. During March – July 2014. students were organized to implement projects of information system development at University of Novi Sad, Technical faculty "Mihajlo Pupin" in Zrenjanin [27]. Overall organization was set according to DAD methodology. Their projects' mentor had a role of stakeholders' representative for each of these projects. Second category of projects did not include business process modeling, but only use case model as well as conceptual and physical data models.

Methods for testing previously represented hypotheses upon the empirical results are related to simple comparison (first type of projects compared to second type of projects) of number/percentage of each indicator's average occurrences, presented graphically. Data in average number/percentage is based on estimation for the category of projects made by mentor of these projects.

In aim to enable comparison, selection of projects for sample is made in aim to have both groups comparable. Both groups of projects show many similarities:

- Both groups of projects have the same number of students (25) working on the same number of projects (20); both groups of students have the same characteristics in both groups: gender (5

female students and 20 male students), study level (10 master-level students and 15 bachelor-level students).

- Mentorship from the same teaching personnel member (i.e. teaching assistant Ljubica Kazi)
- Both groups of software is developed within information systems context, i.e. related to development of software for organizational information systems
- Both groups of projects have equal project average complexity (number of tables from database per project – namely it is 10 database tables per project)

V. CASE STUDY RESULTS AND DISCUSSION

A. Results

Hypothesis 1 related to time indicator – "Business process modeling increase overall project duration".

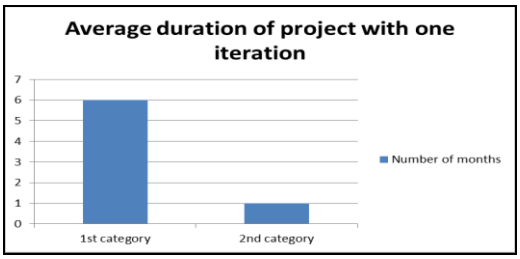


Figure 1. Average duration of project with one iteration comparison

Projects from 1st category average duration is six months including modeling and implementation up to 1st iteration completion, while projects from 2nd category average duration is 1 month.

Hypothesis 2 related to time indicator - "Omitting business process modeling increase number of development iterations in software development".

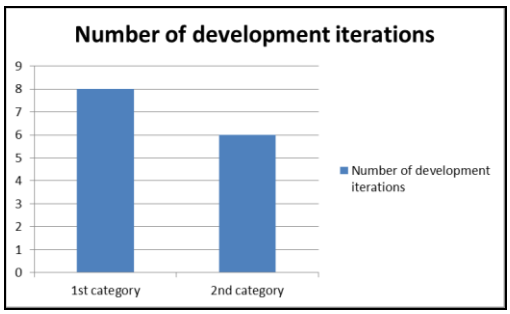


Figure 2. Number of development iterations comparison

Projects from 1st category had average of 8 development iterations, while projects from 2nd category had average of 6 development iterations. In first category, iterations are: business process modeling (1st and 2nd), use case model (1st and 2nd), conceptual data model (1st and 2nd) and software development in first iteration complexity level (1st and 2nd). This average statistics shows that this hypothesis is not valid – statistics proves that 2nd category of projects without business process modeling had less development iterations.

Hypothesis 3 related to scope indicator – "Business process modeling enables completeness of the project scope".

Hypothesis 4 related to scope indicator – "Omitting business process modeling gives partial results of the project scope".

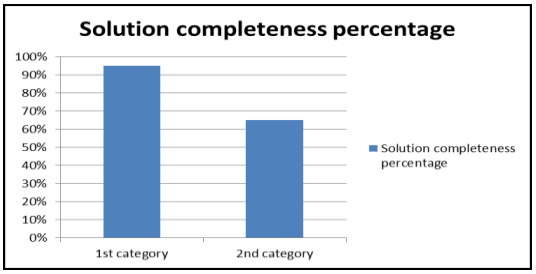


Figure 3. Solution completeness percentage comparison

Percentage of solution completeness is related to the percent of software functions designed compared to the total number of software functions implemented in solution. For the 1st category, percentage of solution completeness designed and implemented is 95%, while in the 2nd category is 65%.

B. Discussion

Case study in this paper is based on estimation of duration of project, number of iterations and percentage of completeness of developed solutions compared to designed features. This estimation is made by projects mentor. Case study is based on comparison of projects from 1st category (mandatory projects of information system development that include business process modeling) and 2nd category (optional projects of information system development that do not include business process modeling).

Regarding first hypothesis, projects from 1st category show much longer duration comparing to projects that do not include business process modeling. Second hypothesis observations show that projects with business process modeling had more development iterations comparing to those that did not have business process modeling included in project activities. Finally, projects that had business process modeling included had more complete final solutions after first iteration, comparing to those having absence of business process modeling.

Finally, it is needed to give answers to previously presented questions:

1st question: *In aim to minimize time and efforts in modeling, could business process modeling be avoided? Could we just do minimal modeling in functional and data segment without previous business process modeling?*

Possible answer: Software solutions within information system development could be developed with minimum modeling including functional aspect (use case) and data aspect (conceptual data model) and business process modeling could be avoided. If time is important, project duration and number of development iterations would be less.

2nd question: What are the consequences if we omit business process modeling?

Possible answer: Omitting business process modeling could lead to incomplete solutions, not having concerned all business process needs and appropriate software mapping. If scope is important, business process modeling brings more complete solution basis.

VI. CONCLUSIONS

This paper aim is to investigate the role of modeling in disciplined agile delivery approach within information system development. DAD methodology denotes "initial modeling", but which level of details is appropriate? In this paper we presented a case study which focused on the role of business process modeling and the need for these models inclusion in DAD approach to information system development. Finally, models needed for DAD approach applied in information systems development should include functional model (such as UML's use case diagram) and data model (conceptual model) as two core types of models. Business process model could be avoided, since it makes project duration longer and increase number of development iterations. Still, business process model inclusion improves project scope, i.e. solution completeness.

This paper conclusions are related to emphasizing the need for functional and data modeling, while business process modeling inclusion in DAD is analyzed from the time and scope perspective. If time perspective is considered more important, business process modeling should be avoided, but if scope perspective is emphasized, business process modeling improve the solution completeness. Finally, there is no unique answer. Business process modeling is a basis for both functional and data modeling, but experienced professionals could include this activity as abstract mapping between business processes to elements of functional and data models, needed for implementation.

REFERENCES

[1] B. Boehm, "A View of 20th and 21st Century Software Engineering", *ICSE'06*, May 20–28, 2006, Shanghai, China.

[2] Agile Software Development Alliance, www.agileAlliance.org [visited 4th January 2015]

[3] Agile Manifesto, http://agilemanifesto.org/ [visited 4th January 2015]

[4] S. W. Ambler: "Going Beyond Scrum, Disciplined Agile Delivery", Disciplined Agile Consortium, White Paper Series, October 2013.

[5] S.W. Ambler and M. Lines, "Disciplined Agile Delivery: A Practitioner's Guide to Agile Software Development in the Enterprise", IBM Press, 2012.

[6] Disciplined Agile Manifesto, https://disciplinedagiledelivery.wordpress.com/disciplinedagilemanifesto/ [visited 5th January 2015]

[7] Lj, Kazi, Z, Kazi, B, Radulovic, "Data Warehouse Based Evaluation of Students' Achievements in Information Systems Education", Proceedings of the IEEE International convention on Information and Communication Technology, Electronics and Microelectronics MIPRO 2012, Opatija, Croatia

[8] Lj, Kazi, Z. Kazi, B. Radulovic, D. Radosav, "Evaluation of Models in Information Systems Development", Proceedings of the International Conference Dependability and Quality Management ICDQM 2011,Belgrade, pp.589-595

[9] Lj. Kazi, Z. Kazi, B. Radulovic, O. Stanciu, "Evaluation of students' work on Data Modeling – Teaching Improvement Implications", *Journal Information technologies and development of education international ITRO*, ISSN 2217-7930, Vol 1, No 1, pp. 24-30, 2011.

[10] Lj. Kazi, Z. Kazi, B. Radulovic, D. Letic, "Using Automated Reasoning System for Data Model Correctness Analysis", Proceedings of the 8th IEEE International Symposium on Intelligent Systems and Informatics SISY 2010, September 10-11 2010, Subotica, Serbia, ISBN 978-1-4244-7394-6, pp. 522-527

[11] Lj. Kazi, B. Radulovic, D. Radosav, M. Bhatt, N. Grmusa, N. Stiklica, "Business Process Model and Elements of Software Design: The Mapping Approach", Proceedings of the International conference on Applied Internet and Information Technologies" ICAIIT2012, Zrenjanin, pp.17-20.

[12] Lj. Kazi, B. Radulovic, I. Berkovic, Z. Kazi, "Integration of conceptual data modeling methods: Higher Education Experiences", Proceedings of IEEE International convention on Information and Communication Technology, Electronics and Microelectronics MIPRO 2014, pp. 963-968.

[13] R. Wendler: "The Structure of Agility from Different Perspectives", Proceedings of the Federated Conference on Computer Science and Information Systems, 2013, pp. 1165–1172

[14] A. Ganguly, R. Nilchiani, and J. V. Farr, "Evaluating agility in corporate enterprises," *International Journal of Production Economics*, vol. 118, no. 2, pp. 410–423, Apr. 2009.

[15] R. Dove, "Knowledge management, response ability, and the agile enterprise," *Journal of Knowledge Management*, vol. 3, no. 1, pp. 18– 35, 1999.

[16] B. Boehm: "Get Ready for Agile Methods, with Care", *IEEE Computer*, January 2002, pp 64-69.

[17] S. W. Ambler, M. Lines: "Disciplined Agile Delivery, The Fonudation for Scaling Agile", *CrossTalk*, November/December 2013

[18] O. Hazzan, Y. Dubinsky: "Why Software Engineering Programs Should Teach Agile Software Development", *ACM SIGSOFT Software Engineering Notes*, Page 1, March 2007, Vol 32, No 2

[19] K. M. Slaten, M. Droujkova, S. B. Berenson, L. Williams, L. Layman: "Undergraduate Student Perceptions of Pair Programming and Agile Software Methodologies: Verifying a Model of Social Interaction", Proceedings of IEEE Agile Conference 2005, 24-29 July 2005, pp. 323 - 330

[20] "Agile Project Management and Software Development" course of Master Studies in Informatics, Technical University München - Fakultät für Informatik, http://www4.in.tum.de/lehre/vorlesungen/vgmse/ws1213/index.shtml [visited 6th January 2015]

[21] D.F.Rico, H.H. Sayani, "Use of Agile Methods in Software Engineering Education", Proceedings of the IEEE Agile Conference, 2009, 24-28 Aug. 2009, Chicago, IL, pp.174 – 179

[22] V. Mahnic, "A Capstone Course on Agile Software Development Using Scrum", *IEEE Transactions on Education*, Vol. 55, No. 1, February 2012

[23] C.G. von Wangenheim, R. Savi, A.F. Borgatto, "SCRUMIA—An educational game for teaching SCRUM in computing courses", *The Journal of Systems and Software,* 86 (2013), pp. 2675 -2687

[24] R.,Elmasri, S.B.Navathe, Fundamentals of Database Systems, 5th edition, Pearson International Edition, 2007.

[25] D.E.Avison, G.Fitzgerald, "Information Systems Development: Methodologies, Techniques and Tools", McGraw Hill, 2003.

[26] "A Guide to the Project Management Body of Knowledge", Fifth Edition, Project Management Institute,2013.

[27] Lj. Kazi, B. Radulovic, M. Ivkovic, B. Markoski, D. Glusac, M. Pavlovic, D. Dobrilovic, D. Lacmanovic, Z. Veljkovic, V. Karuovic:" Improving Information System of Higher Education Institution: a Case Study", Proceedings of International conference Applied Internet and Information Technologies, AIIT2014, October 24,2014, Zrenjanin

[28] Disciplined Agile Delivery - Life Cycle https://disciplinedagiledelivery.wordpress.com/lifecycle/

Domain specific agent-oriented programming language based on the Xtext framework

Dejan Sredojević*, Dušan Okanović**, Milan Vidaković**, Dejan Mitrović***, Mirjana Ivanović***
* Novi Sad Business School, Novi Sad, Serbia
** University of Novi Sad/Faculty of Technical Sciences, Novi Sad, Serbia
*** University of Novi Sad/Faculty of Sciences, Novi Sad, Serbia
dsredojevic.vps@gmail.com, {oki, minja}@uns.ac.rs, {dejan, mira}@dmi.uns.ac.rs

Abstract— **The agent technology represents one of the most consistent approaches to the development of distributed systems. Multiagent middleware XJAF, developed at the University of Novi Sad, presents a runtime environment that supports the execution of software agents. To solve the problem of interoperability, we propose a domain-specific agent language named ALAS, whose main purpose is to support the implementation and execution of agents on heterogenous platforms. To define the structure of the language, a metamodel and a grammar of the ALAS language has been created, in accordance with the requirements and needs of the agents. This paper describes the construction of the compiler and the generation of executable Java code that can be executed in XJAF.**

I. INTRODUCTION

A software agent is a program that works autonomously while performing tasks that are assigned to it [1]. These are target-oriented computer programs which react to their own environment. They work without direct supervision and perform tasks for the end user or other programs. Features of software agents include autonomy, intelligence, mobility, persistence and communication [2]. Autonomy implies that agents must be able to independently perform tasks. Mobility implies that agents have ability to leave the place where they currently execute a task and to continue the execution of the task on another node in the network [3][4]. Communication implies that agents must be able to communicate with other agents in the system.

A system that consists of several software agents is called a *multiagent system* - MAS. Such agents are capable of collectively solving a task that is most difficult to be solved by a single agent or monolithic system. Its main features include agent lifecycle management, messaging, security mechanisms, and service subsystem that gives agents the ability to access resources, execute complex algorithms, etc [1].

The rest of the paper is organized as follows. The *Related work* section describes a couple of existing agent-oriented programming languages (AOPL) that had a strong influence on the development of ALAS. The multiagent middleware *Extensible Java EE-based Agent Framework* (XJAF) [5] is described in the third section. In the fourth section of the paper, an agent-oriented programming language ALAS is described [4][6][7]. The fifth section presents the results of testing the ALAS using the Eclipse framework, along with the Xtext-based plugin installed. Last section gives concluding remarks.

II RELATED WORK

Agent-oriented programming (AOP) [8] is a software development paradigm aimed at efficient development of software agents and multi-agent systems. AOP shares many features with object-oriented programming (OOP), and it is based on agents which definition include agent state, actions, services and messaging system. Since object-oriented and agent-oriented paradigm share many programming concepts, development of an OOP-inspired AOPLs is a natural process.

One of the first AOP languages was AgentSpeak. The AgentSpeak programming language was introduced in [9]. It is a natural extension of logic programming for the beliefs desires-intentions (BDI) agent architecture, and provides an abstract framework for programming BDI agents.

JACK [10] is a light-weight framework for rapid development of multi-agent systems. It is based on the Java programming language, but offers new keywords and language constructs. The accompanying compiler produces pure Java code, which allows for each JACK agent to be used plain Java object.

Agent mobility is an essential property of agents. However, this property can be quite complex to implement. Any AOPL should hide this complexity from end-users. SAFIN [11] and CLAIM [12] hide complex support for agent mobility from the programmer. They hide the functional complexity from developers by providing them with simple, yet powerful programming constructs.

JIAC V [13] is a multi-agent system that can execute agents developed in pure Java or by using an accompanying AOPL named JADL++. Similarly to AgentSpeak, an action of a JADL++ agent can be either private, for internal use, or public, in which case it is called a *service* and offered to other part of the system.

We have introduced the early version of the ALAS programming language in papers [6] and [7]. This early version have used *javacc* [16] parser generator, while this paper proposes the Xtext framework for the language development. We have decided to use Xtext, since we wanted to start with the ECore metamodel, to separate validation from compilation and to have Eclipse plugin made for ALAS without additional programming. This plugin offers both validation and code generation to any programming language.

III XJAF DEVELOPMENT FRAMEWORK

XJAF is a multiagent middleware based on the FIPA standards [14]. FIPA is a non-profit organization which has produced a set of specifications that enable interaction between the agent and the framework, as well as interaction between agents.

The main tasks of XJAF are to provide an efficient environment for the execution of its agents and to provide a reusable interface to external clients [2]. XJAF is designed as a modular system that contains specialised modules called *managers*. Each manager is relatively independent module responsible for handling certain agent management tasks.

The latest version of XJAF is focused on using the advantages of computer clusters [15]:

- Load balancing – XJAF agents are automatically distributed in a cluster in order to reduce the load on individual computer nodes.
- State replication and failover: state of each agent is copied to other nodes making them resistant to hardware and software faults.

The main drawback of the original XJAF architecture was the fact that it was limited to the Java programming language. The consequence of this approach was that the agents needed to be written in Java, and could not interact with agents in other, non-Java frameworks. To increase the interoperability and enable its wider usage, XJAF has been redesigned as a service-oriented architecture - SOA. The multiagent framework based on SOA called SOM - *SOA-based multiagent system*, kept Java EE as the implementation platform, but managers were redefined as web services [3][6]. This enabled that implementation of SOM could be done in many modern programming languages that support web services (Java, JavaScript, Python, C#, etc.). Interoperability is also increased and external clients and independent tools can employ agents through web service interfaces.

However, the use of web services does not solve all the problems. Considering that other programming languages can be used for the implementation of an agent framework, it is impossible to write an agent that will successfully execute on any platform. The problem becomes apparent when the agents that are written in different programming languages, moving through the network arrive to a SOM that is implemented in some other programming language. For example, agent developed using Java programming language can not be easily adapted to agent framework implemented for example, in Python. To solve this problem the authors of XJAF have developed a new agent language - ALAS [7].

IV SPECIFICATION OF ALAS LANGUAGE

The main goals of ALAS are:

- *Hot compilation* - to ensure that the agents can be executed in target platform, regardless on the underlying programming language,
- *Hiding complexity* of agent development from programmers.

Agents must adapt to the environment in which they arrive. When they arrive to some framework that is implemented in a programming language X, they must be automatically transformed to source code written in X.

According to these requirements it is necessary to implement a compiler for the ALAS language. Input parameters for this compiler are the original file written in ALAS, and the identification of the destination platform, such as Java, JavaScript, Python C#, etc. Depending on these parameters, the compiler generates executable code depending on the destination platform. If the platform on which the agent arrived has been implemented in the Java programming language, it will be Java byte code. In this way, developers are able to focus on solving concrete tasks and do not have to take care about interoperability or details of the implementation of SOM.

A. Development of ALAS using Xtext framework

Due to restrictions of the *javacc* system previously used for ALAS transformation [16], development of ALAS has been restarted using Xtext. This framework is most commonly used for domain-specific language development [17].

Xtext enables the development of agent-based domain-specific language ALAS, so it could meet the aforementioned requirements. The development of domain-specific language using Xtext is performed using specialized languages - Java and Xtend. Xtend is a statically-typed programming language which translates to comprehensible Java source code. Syntactically and semantically Xtend has its roots in the Java programming language but it is improved in many aspects. Xtext is an open-source framework for the development of domain-specific languages - DSL. It is based on ANTLR [18]. Unlike standard parser generators, Xtext not only generates a parser, but also a class model for the abstract syntax tree and a fully featured, customizable Eclipse-based IDE.

B. Domain-specific language ALAS modeling

In the last decade, great advancements appeared in the modeling field and defining of domain-specific languages. These languages allow developers and domain experts to focus on specific domain problems instead of dealing with the programming language features. Formation of OMG (Object Management Group), an organization that gathered the most important industrial subjects in order to establish standards in the field, has greatly contributed to the acceleration of the development methodology for designing software controlled by models [19]. OMG initiative called MDA (Model-Driven Architecture) represents one of the most important movements in the development of software controlled by models. MDA is a specialization of a broader approach called MDE (Model-Driven Engineering) which is a development metodology focused on creating domain models. The domain-specific languages are programming languages designed for solving problems in a specific, clearly defined, domain.

MOF (Meta-Object Facility) is an OMG standard, which represents the core of infrastructure for support to MDA [20]. MOF is a specification that enables

cooperation between different domains and different modeling language and is a mechanism for formally defining modeling languages, i.e. metamodels. MOF is a four layered architecture whose layers are marked as M0, M1, M2 and M3 (Fig. 1). MOF can be viewed as an abstract syntax of DSL which used to define the metamodel.

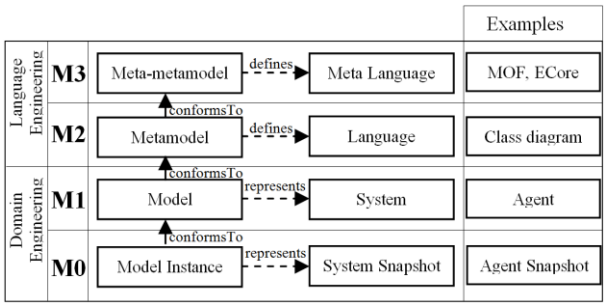


Figure 1. Four-layered metamodel architecture

There are several implementations of MOF infrastructure which more or less follow specification standards. Implementation that is used to create a metamodel of ALAS is ECore. ECore was developed by the Eclipse Foundation and EMF project, which basically is an implementation of EMOF 2.0. ECore is also open source and is part of the Eclipse platform [21].

C. *Implementation of ALAS metamodels, by using ECore meta-metamodel*

To define domain-specific language ALAS, a metamodel has been created first i.e. class diagram of our domain-specific language. ALAS metamodel is implemented by using ECore meta-metamodel which is a part of Eclipse framework.

To represent the metamodel graphically, the Eclipse framework uses Sirius 2.0 plugin [22]. ALAS ECore metamodel may be defined manually, using the tree editor, but this kind of modeling is tiresome and impractical. Other, often used method is to define the metamodel using the class diagram which is accessible in existing modeling tools. In Fig. 2 you can see the class diagram, which defines the structure of the domain-specific language ALAS.

D. *Grammar of ALAS language*

Created metamodel is then used for automatic generation of the ALAS grammar. After the integration of Xtext plugin into the Eclipse framework, it is possible to create an Xtext project which will implement the previously mentioned ECore metamodel. The grammar of ALAS language is shown in Listing 1. Considering that the graphic representation of the ALAS metamodel does not fully define the structure of the language, it is necessary to modify the generated grammar by adding keywords and identifiers that are listed under the single quotes and new rules which can be created only after import of certain packages.

Considering that Xtext is compatible with Java programming language, existing Java based rules such as: `XimportSection`, `XblockExpression`, `XExpression`, `XMoveExpression`, `XprimaryExpression` and `XvariableDeclaration` can be usedafter import package from Listing 1, line 4. These rules are automatically added to ECore metamodel of the language and marked as red frame in Fig. 2.

Each rule in the grammar represents an appropriate class in the ALAS metamodel diagram. The grammar contains a set of declared events, variables, functions, services, actions and agent states. The grammar is based on the EBNF - Extended Backus-Naur Form - Listing 1.

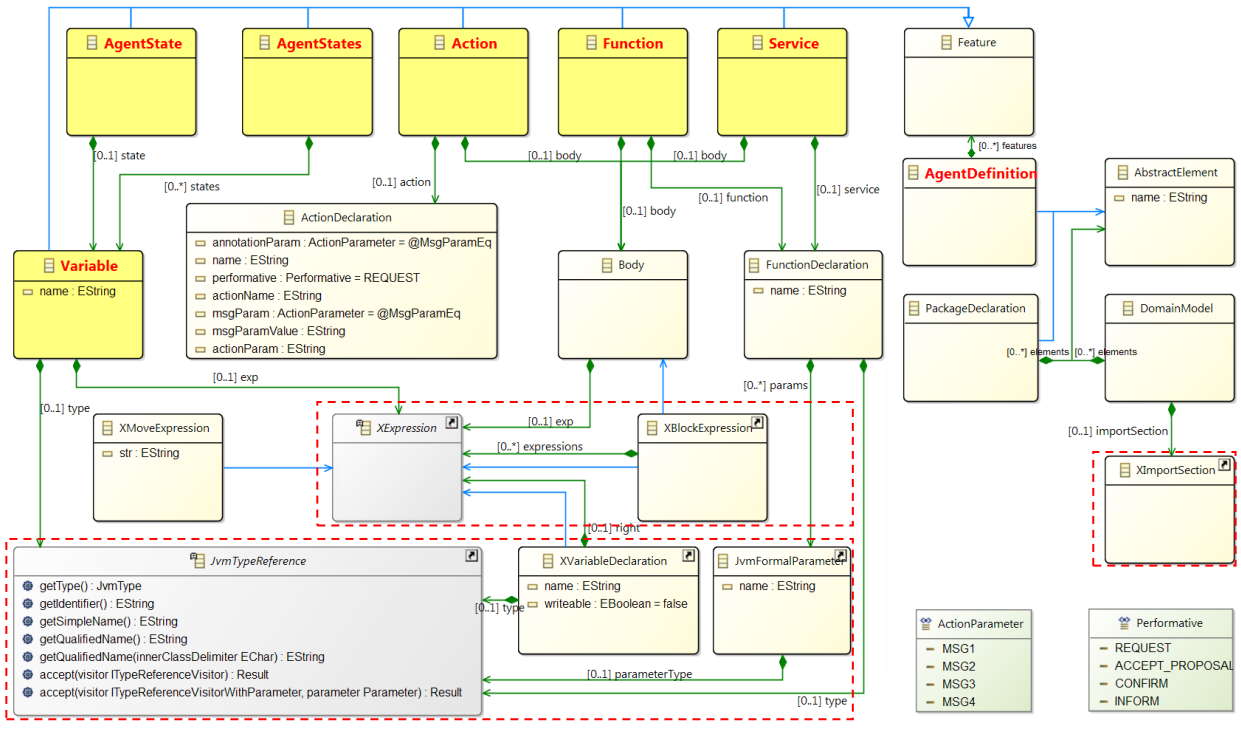


Figure 2. ALAS metamodel

```
1.  grammar rs.ac.uns.alas.Alas with
org.eclipse.xtext.xbase.Xbase
2.
3.  generate alas
"http://www.ac.uns/rs/alas/Alas"
4.  import
"http://www.eclipse.org/xtext/xbase/Xbase"
5.
6.  DomainModel:
7.      importSection = XImportSection?
8.      Elements += AbstractElement* ;
9.
10. AbstractElement:
11.     PackageDeclaration | AgentDefinition ;
12.
13. PackageDeclaration:
14.     'package' name = QualifiedName '{'
15.     Elements += AbstractElement*
16.     '}' ;
17.
18. AgentDefinition:
19      'agent' name = ValidID '{' features +=
         Feature* '}' ;
20.
21. Feature:
22.  =>AgentStates | AgentState | =>Variable |
       Function | Service | Action;
23.
24. AgentStates:{AgentStates}
25.     'state' '{'states += Variable* '}' ;
26.
27. AgentState:
28.     'state' state = Variable ;
29.
30. Variable:
Type = JvmTypeReference name = ValidID ( '='
exp = XExpression)?';' ;
31.
32. Function:
33.     Function = FunctionDeclaration body =
Body ;
34.
35. Service:
36.     'service' service =
FunctionDeclaration body = Body ;
37.
38. Action:
39.     Action = ActionDeclaration body = Body
;
40.
41. FunctionDeclaration:
43.     type = JvmTypeReference name=ValidID
44.     '('(params += FullJvmFormalParameter
(',' params += FullJvmFormalParameter)*)?')' ;
45.
46. XMoveExpression returns XExpression:
47.     {XMoveExpression}'move'
       '('(str=STRING)')' ';' ;
```

Listing 1. Part of ALAS grammar

After any change in the language grammar, it is necessary to generate the appropriate artifacts, i.e. language infrastructure.

The first rule in the grammar - `DomainModel` is always used as an input or initial rule. In this case, the

```
1. int value = 1;
2. String str = "host";
```

Listing 2. An example of variable definitions by using 'Variable' rule

`DomainModel` may or may not (quantifier ?) contain an imports. Also, it contains an arbitrary number (quantifier *) of `AbstractElement` rules that will be added to the parameter elements (quantifier **+=**). Within the `AbstractElement` rule the `PackageDeclaration` or `AgentDefinition` rule (quantifier '|') can be used. Within the `PackageDeclaration` rule the name of the package is defined the program will be written and again within the same package – `PackageDeclaration`, the choice can be made between `PackageDeclaration` rule or `AgentDefinition` rule.

`AgentDefinition` rule defines an agent. Agent is defined with keyword *`agent`*, then his name, and then the body of an agent in brackets ('{' and '}'). Agent can contain an arbitrary number of `Feature` rules. The `Feature` rule allows us to write any of the following six rules: `AgentStates`, `AgentState`, `Variable`, `Function`, `Service` or `Action`. Some of these rules can be added within another, and this can lead to compiler error, since the compiler does not know which rule to execute. To prevent this, we use the '**=>**' quantifier. It gives an advantage to the rule, in front of which is located. This way the compiler first applies this rule.

The difference between the first two rules, `AgentStates` and `AgentState` is as follows: within the `AgentStates` rule an arbitrary number of `Variable` rules can be defined, and within the `AgentState` rule only one `Variable` rule can be defined. This definition enables us to define a variable in the form in Listing 2.

One of the future goals of the ALAS programming language will be to provide mobility of an agent. One of the functions that will be used for this purpose is the *`'move'`* command - `XMoveExpression`. This command is specified in grammar of the language and is implemented within the `XblockExpresion`. To achieve this, the `XPrimaryExpression` rule has been redefined from the source of the Xbase grammar by adding the `XMoveExpression` rule. The `XMoveExpression` rule introduces a new keyword *`'move'`* and within the brackets that follow a String argument is placed. This argument represents an address to which the agent shall move (Listing 4, line 32).

Because the agent is written in ALAS, regardless of the task or problem that it solves, it can not be used as such in some environment that uses a different programming language. This is the appropriate moment to introduce the conversion - mapping of an agent to the destination programming language. The next section will describe the process of generating Java code from a program written in ALAS.

E. Translating program from ALAS agent language to Java code

To generate code for target platform, from an agent written in ALAS language, we can use automatically generated class that is provided by Xtext – it is generated with other language infrastructure constructs. The name of the used grammar must be provided at the beginning of the grammar file. ALAS uses the `Xbase.xtext` grammar, which can be seen in the line 1 of Listing 1. Based on the built-in grammar that is used, the compiler generates packages and classes that will be used for mapping to the concrete programming language. If the grammar uses `Terminals.xtext`, it will generate the package *generator* and within it classes that will be used

for mapping to different programming languages. Since the Xtext is closely related to the Java programming language and ALAS is Java-like language, the `Xbase.xtext` grammar is used. The parser will not generate the *generator* package but *jvmmodel* package and within it `AlasJvmModelInferrer` class that will be used to translate an agent written in ALAS to the Java program.

The `AlasJvmModelInferrer` class is implemented in the Xtend programming language. The main goal of the ALAS is that it can be transformed to any programming language, not only to Java. To achieve this, it was necessary to reimplement the `AlasJvmModelInferrer` class with all the rules that will be used by the parser to generate any destination code. To do so it was necessary to implement the `AlasModelGenerator` class and the `doGenerate` method, which generates destination code using rules from the `AlasJvmModelInferrer` class (Fig. 3).

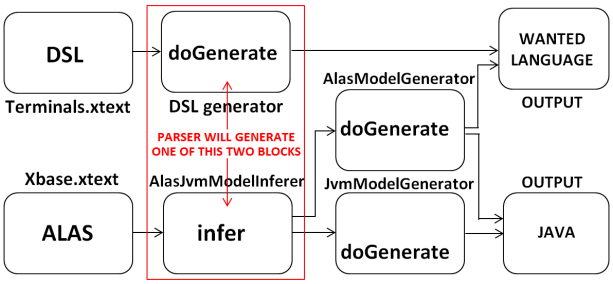


Figure 3. Working principle of parser generators

To implement a *service*, rules in the Listing 3 were used. Listing 4, lines 14 to 33, shows a service written in the ALAS language, while Listing 5, lines 18-28, shows Java implementation of that service. The Java code was generated from code in the Listing 4, applying the rules from the Listing 3. The *parameter* keyword was used to define *service* parameters.

```
1. Service : {
2.    members += f.toMethod(f.service.name,
         f.service.type ?: inferredType) [
3.      documentation = f.documentation
4.      visibility = JvmVisibility.PRIVATE
5.      for (p : f.service.params) {
6.         parameters +=
      p.toParameter(p.name, p.parameterType)
7.      }
8.      body = f.body.exp
9.    ]
10.}
```

Listing 3. The rules for mapping services from ALAS to Java

Considering that agents are Java-like and that writing the body of services, functions and actions uses `XblockExpression`, the parser easily generates the body by applying the *body* rule. If the target language is not Java, the process of generating code is more complex and mapping is then performed in the `AlasModelGenerator` class.

F. Agents code validation

One of the main advantages of a domain-specific language is the possibility of static validation of code segments. Xtext provides the ability to define validation rules and constraints. During the generation process an

```
1. package example.agents{
2.
3.    agent TimeSync {
4.
5.      state {
6.        String startingHome;
7.        String next;
8.      }
9.      state String remaining;
10.
11.     String host = "192.168.0.1";
12.     int n;
13.
14.     service void syncTimers (String hosts,
double time){
15.       if(startingHome == null){
16.         startingHome = host;
17.       }
18.       else if (startingHome.equals(host)){
19.         System.out.println("I'm back home");
20.      startingHome = null;
21.       }
22.       //apply the time
23.         print("Setting the system time to "+
time);
24.         print("ALAS command:
applySystemTime(time)");
25.
26.       //go to the next host
27.       if(hosts.length() == 0) {
28.         next = startingHome;
29.       }
30.       else
31.         parseHosts(hosts);
32.         move("192.168.0.2");
33.     }...
```

Listing 4. Part of test.alas

appropriate *validator* package is automatically generated. This package contains an implementation of the validation rules and required constraint definitions.

The `AlasValidator` class implements various constraints: checking the validity of local and global variables, checking arguments and names of functions, services and activities. In order to invoke validation, the `@Check` annotation has been introduced. This annotation triggers the validation process. An example of validation triggering is shown in the Fig. 4.

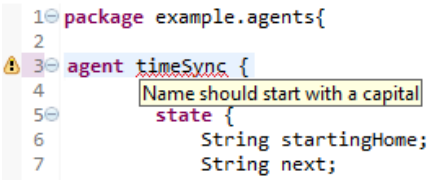


Figure 4. Example of Validation

V TESTING AND RESULTS

The framework is tested using the Eclipse framework with the Xtext-based plugin installed. This plugin enables users to create a new project with the ALAS source code in it - `test.alas` file in Listing 4. The code is automatically converted to Java code and part of the resulting code is displayed in Listing 5.

One of the requirements of XJAF framework is that generated code should contain *onMessage* method - a part of the generated *onMessage* method is shown in Listing 5,

```
1. @Stateful
2. @Remote(AgentI.class)
3. @SuppressWarnings("all")
4. public class TimeSync {
5. @AgentState
6. private String startingHome;
7.
8. @AgentState
9. private String next;
10.
11. @AgentState
12. private String remaining;
13.
14. private String host = "192.168.0.1";
15.
16. private int n;
17.
18. private void syncTimers(final String hosts,
    final double time) {
19.   boolean _equals =
Objects.equal(this.startingHome, null);
20.   if (_equals) {
21.     this.startingHome = this.host;
22.   } else {
23.     boolean _equals_1 =
this.startingHome.equals(this.host);
24.     if (_equals_1) {
25.       System.out.println("I\'m back home");
26.       this.startingHome = null;
27.     }
28.   }
     //..
29. public void onMessage(final ACLMessage msg)
    {
30.   if(msg.getPerformative() ==
31.     Performative.REQUEST){
32.     String s = msg.getContent().toString();
33.     JSONParser parser = new JSONParser();
34.     JSONObject json;
35.     try {
36.       json = (JSONObject) parser.parse(s);
37.       String serviceName =
            json.get("serviceName").toString();
38.       switch (serviceName){
39.         case "syncTimers":
40.          String hosts =
            json.get("hosts").toString();
41.          double time =
Double.parseDouble(json.get("time").toString())
42.
43.            syncTimers(hosts, time);
44.     }//..
```

Listing 5. Part of TimeSync.java

lines 29 to 44. Within the *onMessage* method the data is received in the JSON format [23]. JSON or JavaScript Object Notation, is an open standard format that uses human-readable text to transmit data objects consisting of attribute–value pairs. All the necessary JSON-related libraries must be imported and invoked. By specifying the *onMessage* method, parser automatically imports all relevant classes in the executable Java class.

VI CONCLUSION

Based on the previous work on agent domain-specific language ALAS [4][6][7], we can conclude that the Xtext is the favorable environment for the development of domain-specific languages like ALAS because not only the syntax and validation can be defined, but even the code in an arbitrary target programming language can be generated. By using the Xtext framework it is possible to write agents which will execute specific tasks, but which will also be automatically converted to the executable code of the target platform. Because of these advantages Xtext framework is a better tool than the *javacc* which was used in the previous version of ALAS [7]. This tool can be used to generate source code for other agent-oriented languages, as well as general purpose languages used for agent-oriented programming [24].

Future work for the ALAS language will include implementation of the `AlasModelGenerator` class which will be able to transform ALAS code to an arbitrary programming language. So far, we have been able to manually transform ALAS code to JavaScript and Python, so the `AlasModelGenerator` will be implemented to transform ALAS code to any of those two programming languages. The future work will include analysis of suitability of other programming languages for the XJAF framework. Finally, it is necessary to integrate the Xtext-based plugin into the XJAF framework, so the ALAS code transformation could be performed outside the Eclipse framework. That way, agents written in ALAS will be able to migrate between different XJAF servers.

VII REFERENCES

[1] Vidaković, M., Ivanović, M., Mitrović, D., Budimac, Z.: Extensible Java EE-based agent framework - past, present, future. In: Ganzha, M., Jain, L.C. (eds.) Multiagent Systems and Applications, Intelligent Systems Reference Library, vol. 45, pp. 55 - 88. Springer Berlin Heidelberg, 2013

[2] M. Ivanović, M. Vidakovi´c, D. Mitrović, and Z. Budimac. Evolution of Extensible Java EE-Based Agent Framework. In G. Jezić, M. Kusek, N.-T. Nguyen, R. Howlett, and L. Jain, editors, Agent and Multi-Agent Systems. Technologies and Applications, volume 7327 of Lecture Notes in Computer Science, pages 444–453. Springer Berlin Heidelberg, 2012.

[3] Mitrović, D., Ivanović, M., Budimac, Z., Vidaković, M., - An overview of agent mobility in heterogeneous environments, Proceedings of the workshop on Applications of Software Agents, pp. 52 – 58, 2011

[4] Mitrović, D., Ivanović, M., Budimac, Z., Vidaković, M., —Supporting heterogeneous agent mobility with ALAS, ComSIS, Vol. 9, No. 3, pp. 1203-1229, 2012

[5] Bădică, C., Budimac, Z., Z., Burkhard, Hans-Dieter, and Ivanović, M., „Software Agents: Languages, Tools, Platforms“, Computer Science and Information Systems, ComSIS 8(2), pp. 255–296, 2011

[6] Mitrović, D., Ivanović, M., Vidaković, M., „Introducing ALAS: a novel agent-oriented programming language“, In Symposium on computer languages, implementations and tools, SCLIT 2011, Greece, pp. 19–25, 2011

[7] Mitrović, D., Ivanović, M., Vidaković, M., Sredojević, D., “Okanović, D., Integracija agentskog jezika ALAS u Java agentsko okruženje XJAF“, XX naučna i biznis konferencija, 9-13. Mart 2014. pp. 457-461, 2014

[8] Shoham, Y., —“Agent-oriented programming“, Robotics Laboratory Computer Science Department, Stanford University Stanford, CA 94305, USA, 1993

[9] Anand S. Rao, "AgentSpeak(L): BDI agents speak out in a logical computable language", Springer Berlin Heidelberg, number 1038, pp 42–55, 1996

[10] M. Winikoff, "Jack™ Intelligent Agents: An Industrial Strength Platform," in Multi-Agent Programming: Languages, Tools and Applications, Springer US, pp 175-193, 2005

[11] D. Xu, G. Zheng, and X. Fan, "Information and Software technology", pp. 435-442, 1998

[12] A. E. Fallah-Seghrouchni, and A.Suna, "CLAIM and SyMPA: A Programming Enwironment for Intelligent and Mobile Agents," in Multi-Agent Programming: Languages, Tools and Applications, Springer US, pp. 95-122, 2005

[13] B. Hirsch, T. Konnerth, and A. Heßler, "Merging Agents and Services – the JIAC Agent Platform," in Multi-Agent Programming: Languages, Tools and Applications, Springer US, pp. 159-185, 2009

[14] FIPA Abstract Architecture Specification, http://www.fipa.org/specs/fipa00001/SC00001L.pdf, 12.10.2014.

[15] Mitrović, D., Ivanović, M., Vidaković, M., Budimac, Z., Extensible Java EE-based Agent Framework in Clustered Environments, 12th German Conference, MATES 2014, 23-25. Septembar, Štutgart, Nemačka, Proceedings, pp. 202-215, 2014

[16] JavaCC, https://javacc.java.net/ 15.11.201.

[17] Xtext, http://www.eclipse.org/Xtext/documentation.html, 12.10.2014.

[18] ANTLR, http://www.antlr.org, 15.10.2014.

[19] Object Managament Group, http://www.omg.org 15.11.2014.

[20] Meta Object Facility, http://www.omg.org/mof 15.11.2014.

[21] Eclipse Modeling Framewok, http://www.eclipse.org/modeling/emf 15.11.2014.

[22] Sirius 2.0, http://eclipse.org/sirius/ 20.11.2014.

[23] JSON http://docs.oracle.com/javaee/7/tutorial/doc/jsonp.htm, 18.10.2014

[24] Pokahr, A., Braubach, L., Haubeck, C., Ladiges, J., "Programming BDI agents with pure java," In: Muller, J.P., Weyrich, M., Bazzan, A.L. (eds.) Multiagent System Technologies, Lecture Notes in Computer Science, vol. 8732, pp. 216–233. Springer International Publishing, 2014

Aspect-Oriented Engines for Kroki Models Execution

Milorad Filipović, Sebastijan Kaplar, Renata Vaderna, Željko Ivković, Gordana Milosavljević, Igor Dejanović

Faculty of Technical Sciences, Novi Sad, Serbia
{mfili, skaplar, vrenata, zeljkoi, grist, igord}@uns.ac.rs

***Abstract* – The paper presents an overview of techniques and mechanisms implemented in generic web and desktop engines that enable execution of application prototypes being specified by our Kroki tool. Unlike most other solutions where only a user interface skeleton is executable, Kroki's specifications can be tested through all three application tiers – the user interface, the business logic, and the database. Kroki is a mockup-driven tool that enables development of enterprise information systems based on participatory design. Since immediate execution is always possible, it can significantly contribute to decreasing a communication gap between the development team and users.**

I. INTRODUCTION

Kroki [1, 17, 18] is a rapid prototyping tool that enables users and developers to be concurrently engaged in the development of enterprise information systems. Kroki enables requirements elicitation based on executable prototypes, using the terms that are familiar to the end users - by enabling them to draw the user interface (UI) mockups. Contrary to the approaches where mockups are created by general-purpose drawing tools and then manually or semi-automatically transformed to formal models (which are prone to errors and can lead to information loss), mockups created by Kroki are already elements of the UI model. Kroki's mockup editor actually implements a concrete syntax of our EUIS (Enterprise User Interface Specification) DSL [13] for specifying UIs of enterprise applications at a high-level of abstraction. EUIS DSL also has a textual syntax implemented by Kroki's command window and a UML-like concrete syntax implemented by Kroki's UML lightweight editor (Figure 5) [18]. EUIS DSL supports specification of several types of forms and panels and their elements, where the corresponding layout and functionality are defined by our user interface guidelines.

In order to reduce waste of time and effort, a special attention is paid to the option of reusing artifacts across development phases. The reuse is supported by exporting class diagrams and application prototypes to general purpose modeling and programming tools, and by importing models from general-purpose modeling tools (Figure 1). Thus, a created prototype can be used for requirements elicitation and can also evolve to the final enterprise application using the preferred toolchain (currently supported target language is Java).

Kroki enables hands-on prototype evaluation based on executable engines that can be activated almost instantaneously at any given moment during the development phase. This helps narrow the gap between the user specification and the finished product by iterative and online evaluation based on a real working system. According to the principles of

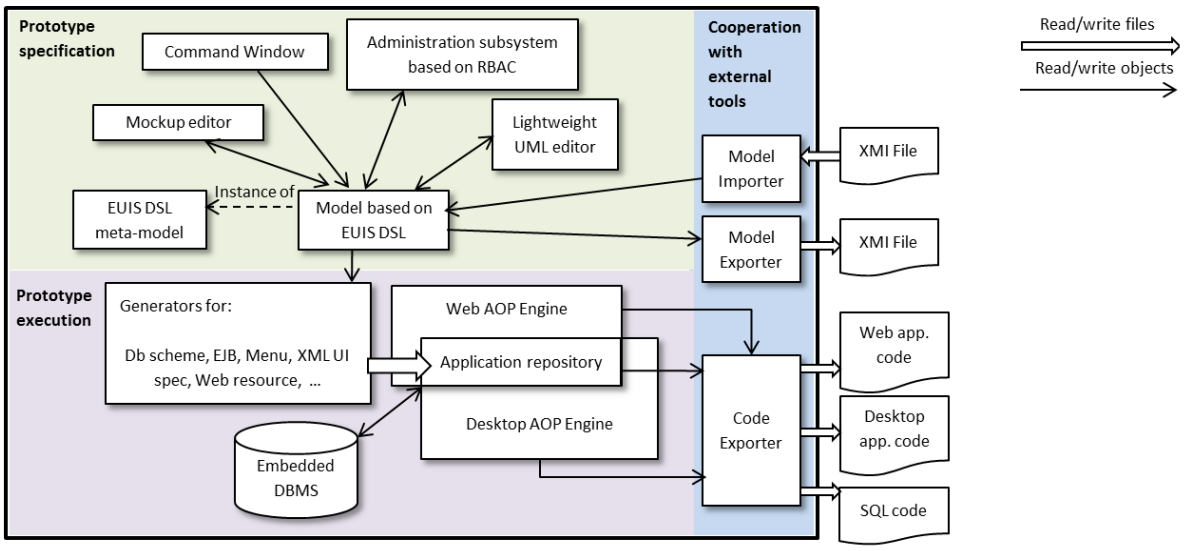


Figure 1. Kroki tool architecture

agile development, information gathering is most effective if it is based on something that works. Unlike most other solutions where only the UI skeleton is executable, Kroki's executable mockups can be tested through all three application tiers – the user interface, the business logic, and the database.

Generic enterprise engines can adapt their functionalities to each specified prototype on the fly. The basis of this adaptable behavior is the configuration data stored in the application repository (Figure 2) [14]. The application repository is a file directory that contains configuration files that provide information about the developed prototype that the generic engines need in order to obtain functions and look defined in Kroki editors. When the user chooses to execute the specified prototype, Kroki generates an application repository specific to the prototype and runs desired (web or desktop) generic enterprise application.

Figure 1 shows the architecture of the Kroki tool. As can be seen, two main parts of the architecture are the Prototype specification and the Prototype execution modules. The paper gives an overview of the prototype executions modules, primarily focusing on the web AOP engine and the application repository. More details about Kroki architecture can be found in [1].

The paper is structured as follows. Section 2 reviews the related work. Section 3 provides a detailed overview of Kroki's application repository with its static and generated parts. Basic mechanisms and principles of generic web engine used for prototype execution are given in Section 4. Section 5 provides additional options for extending built-in engine functionalities. Section 6 gives some final thoughts on the subject of the paper.

II. Related Work

This review of the related work deals primarily with the problems of mockup-driven development and applying aspect-oriented programming in web development.

The generic nature of the developed engine along with its need for adaptiveness leads to a lot of design challenges that make the standard object-oriented and model-driven approaches insufficient and error prone [3]. The shortcomings of traditional approaches are especially emphasized in the design of modern enterprise web applications which are expected to provide a rich user interface and high performance by default [3, 5]. In order to challenge those problems, aspect-oriented programming methods are incorporated into web application development more often than before [4, 5].

The benefits of AOP approach can be seen in [3] and [4], where the greater attention has been dedicated to design techniques of adaptive, context-aware web applications and the performance of the final product. Our approach, which is presented in this paper, follows these basic principles, but presents AOP web engine as a basis for model execution.

The main motivation for developing a generic web engine (in contrast to the standard methods which rely on code generators) was to increase interest in

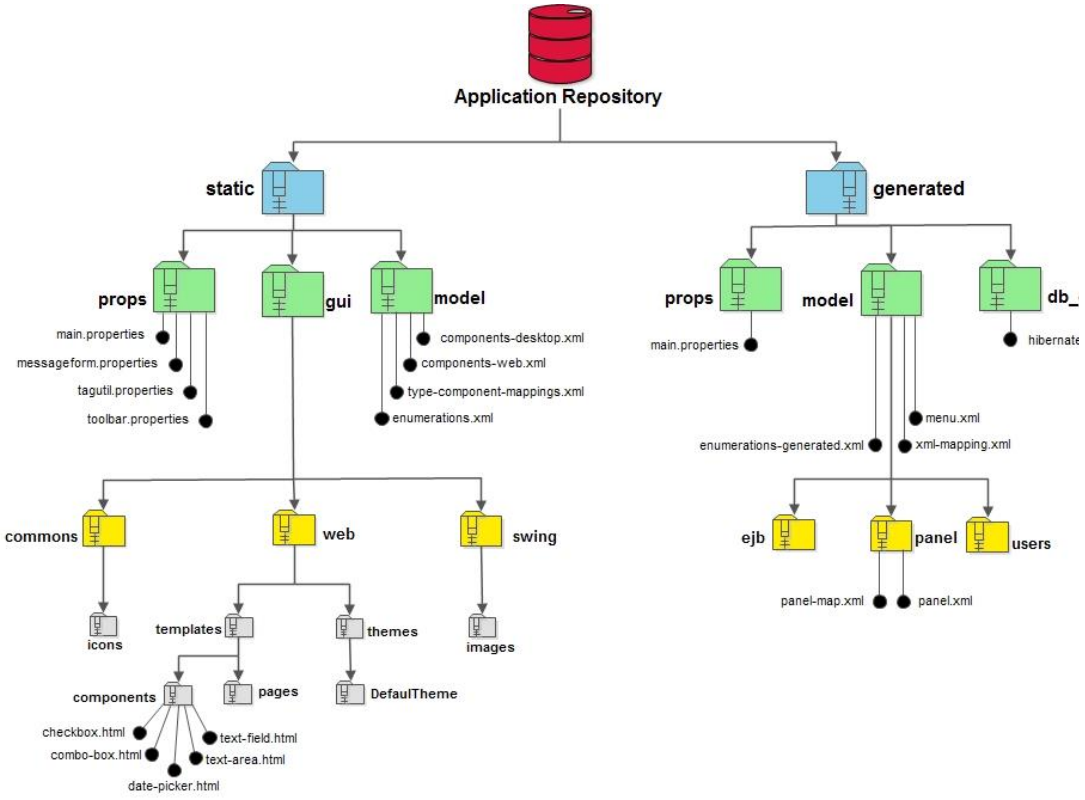


Figure 2. Application repository structure

adoption of agile development methods in web design as presented in [4, 5]. One of the most comprehensive solutions in the field of prototype-driven development is Umple tool, most notably its UIGU generation extension [7]. A slightly different approach is presented in [8] which presents window/event diagrams (WED) as reusable specification artifacts. WEDs combine UI mockups and state diagrams that enrich the UI specification with a prototype which has some basic functionalities. A large body of research deals with digitalization of hand-drawn UI mockups using shape recognition algorithms [8, 9, 10, 11, 12]. Unlike some of the solutions presented here where only the user interface skeleton is executable, Kroki's executable mockups can be tested through all three application tiers.

III. APPLICATION REPOSITORY

The application repository stores configuration data which is the basis for adaptive behavior of Kroki's generic engines. These data specify various parts of the enterprise system and their relations, look-and-feel resources (graphic icons, CSS, and HTML templates etc.), and other configuration artifacts needed for configuration of all generic engine layers.

The application repository structure is shown in Figure 2. It is composed of static and generated parts.

A static part of the repository contains general data that is independent of the concrete specification and as such is always the same (configuration files needed for engine core functionalities, look-and-feel artifacts for web and desktop engines, etc.). Main directories in this part of the repository are:

- **props** - Contains properties files with settings and string resources for web and desktop applications.
- **model** - Contains XML specifications of static parts of the engines. These static specifications mainly deal with the process of mapping programming language types to concrete GUI components
- **gui** - Contains look-and-feel resources for both web and desktop generic applications.

A generated part of the repository is created by Kroki generators and contains data about currently specified application prototype. Although the engine could take this data directly from Kroki model, we choose XML files as an intermediate step in order to provide independent functioning of the specified applications (after deployment). It's structure resembles the structure of the static part. The main difference is the lack of gui subdirectory which is due to enterprise systems using the same UI guidelines used as a basis for EUIS DSL specification [13]. Apart from the configuration files, the concrete EJB classes are being generated directly to the engine source code directory. Hence, they are not part of the application repository.

Main subdirectories of this part of application repository are:

- **props** - provides additional information that supplements the static properties with the data specific to the current specification (such as application name and description).
- **db_config** - contains the hibernate.cfg.ml file used by generic engines to configure the database connection used in the prototype execution phase. During the specification, each engine can be configured to use an existing database or to run embedded test database.
- **model** - contains XML descriptions of enterprise application elements organized in the following files and subdirectories:
    - **ejb** - contains XML specifications of EJB entities used in Kroki project. One XML file is generated for each EJB entity.
    - **panel** - contains XML definitions of standardized panels specified in Kroki project and mapping information (with which EJB entities the panel is associated with)
    - **enumerations-generated.xml** - XML specification of enumerations specified in the Kroki tool.
    - **menu.xml** - XML specification of the application's main menu. The structure of this menu reflects the structure of the packages and forms contained by the Kroki project, but can be overridden by Kroki's administration subsystem. Every user group can have their own main menu.
    - **xml-mapping.xml** - specifies which EJB class is associated with which XML description file in ejb subdirectory.
- **users** - contains XML description files for user rights administration module.

IV. GENERIC WEB ENGINE

Kroki web engine is a generic web application developed in Java that adapts its look and behavior according to the configuration data stored in the application repository. This section presents its architecture in order to provide detailed insight into the engine's inner mechanisms.

Figure 3 shows conceptual architecture of the web engine. Upper half of the figure shows basic modules for initial data collection which are used by both the desktop and web engine. Lower part (with a blue background) displays web-specific architecture. As can be seen, the two parts are loosely coupled, so just the presentation layer is technology dependent. This section will cover some of the basic mechanism for obtaining executable prototype from Kroki

specification with the focus on the web engine. This presented explanation features as less as possible web-specific details, so it can be used to comprehend also Kroki's generic desktop engine since it lies on the same foundations.

is equivalent to the main form with the main menu in the desktop enterprise applications. It shows a page with a main menu from which the user can activate desired form associated with a specific enterprise entity.

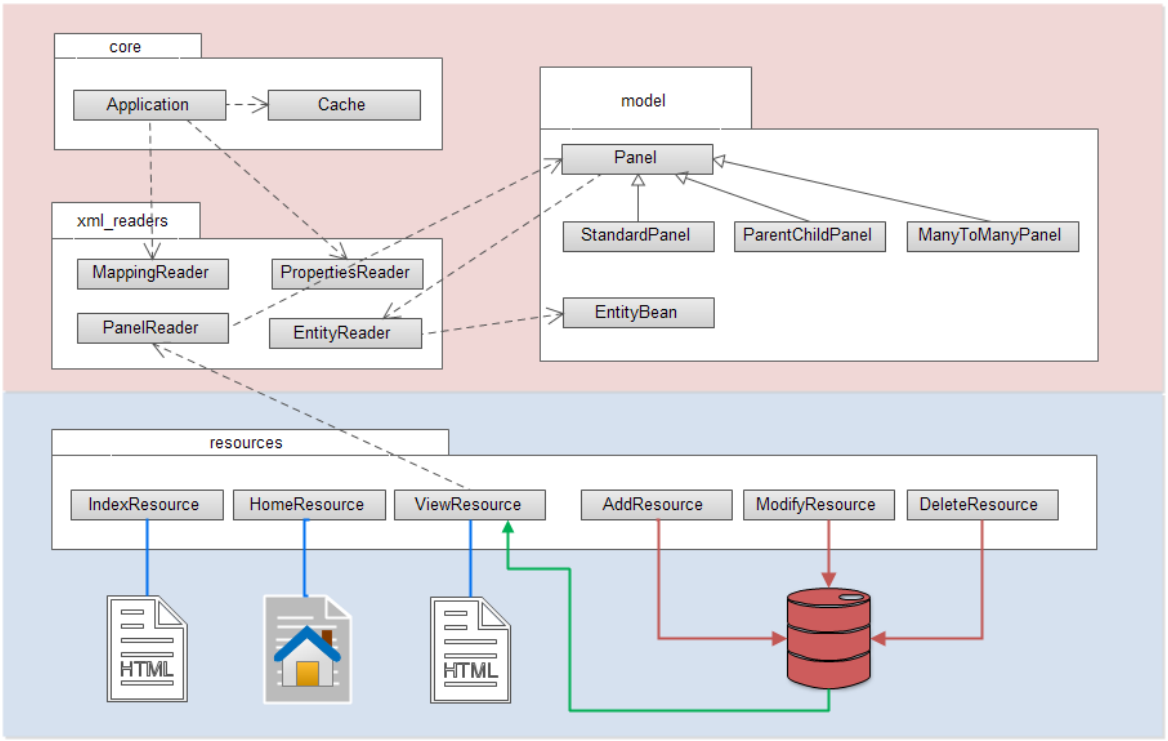


Figure 3. Generic WEB engine architecture

The main components of the engine are shown in the core package in Figure 3. Application module represents the main application class while all of the application's data is stored and managed by the Cache class. Upon startup, application loads the mapping data and project properties from the application repository using the corresponding readers from the xml_readers package. This reduces performance drops when executing large projects since only the mapping data is loaded into an application cache while the actual model data is loaded on demand. Each reader module reads the data from the corresponding configuration file stored in the application repository and stores it in the application cache. Once all the necessary data has been obtained, engine is ready to run.

Basic use scenario in enterprise system revolves around users manipulating data from the database via standard forms. Standard forms contain (one or more) standard panels with well-defined look and features (see [13] for details). The resources package contains the modules responsible for obtaining data for a specific standard form and presenting it to the user. Since our web engine is based on the Restlet web engine, all modules represent Restlet's resource classes. HomeResource handles the login requests and

These user actions are handled by the ViewResource module. It obtains corresponding panel data from the application repository via the PanelReader component. The panel specification contains only the representational aspect of one panel (layout specification and default panel controls), so in order for the given form to be functional, additional persistent data needs to be acquired. Each panel is associated with one EntityBean instance that it obtains via the EntityReader module. Combining the EntityBean and Panel data, the ViewResource displays the web form that conforms to the desired specification. The data that needs to be represented is wrapped into HTML elements using the Freemarker templates.

The basic steps in this process are illustrated in Figure 4. The corresponding Restlet resources handle other enterprise operations. For the sake of simplicity, Figure 3 only shows the basic CRUD (create, update, delete) resources. These modules don't have the explicit HTML representation, they just inform the user of the operation result via the simple text sent over an AJAX call.

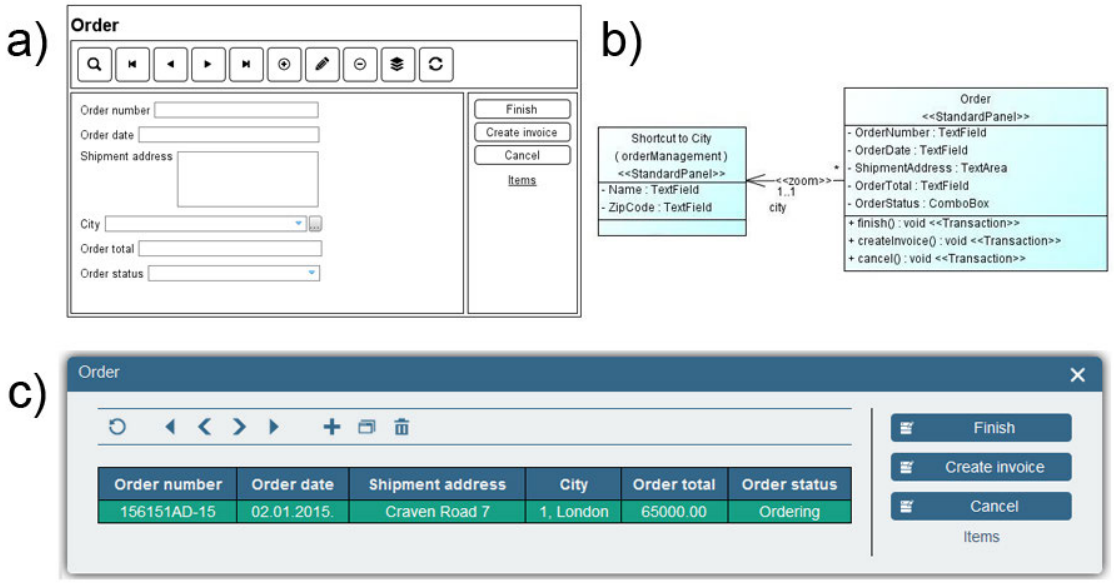


Figure 5 Example of a) a mockup specification b) a corresponding UML-like specification c) a resulting web form in a view mode

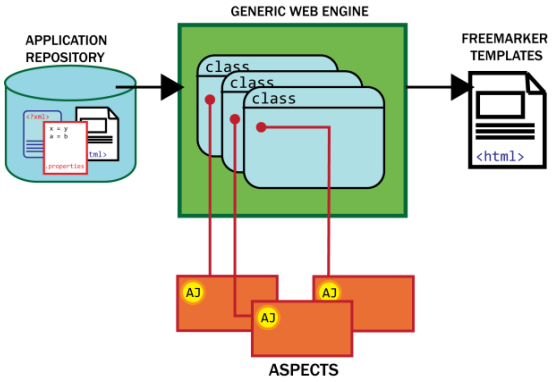


Figure 4. WEB engine overview

V. EXTENDING GENERIC ENGINE

Kroki engines offer standard enterprise functionalities over arbitrary data sets, so its main concern is to process and present data from the database in the predefined way. As a result of the fact that enterprise systems vary in their functionalities, it was necessary to develop mechanism for extending generic engines. Kroki generic engines use aspect-oriented programming techniques to capture run-time points of interests and react in a desired way.

In the web engine, this process is pretty straightforward. The web engine is developed using Restlet engine, so all of the web classes extend Restlet Resource class and are located in the resources package. Every resource class has `prepareContent` method that is invoked when a client request is sent to a particular resource and can be used to attach aspect functionalities. Restlet resources use map called `dataModel` to pass arbitrary data to HTML templates, so once attached to prepareContent, aspect can get access to the resource object and its corresponding `dataModel` (Figure 6). Also, all data contained in the application cache is available to the aspect.

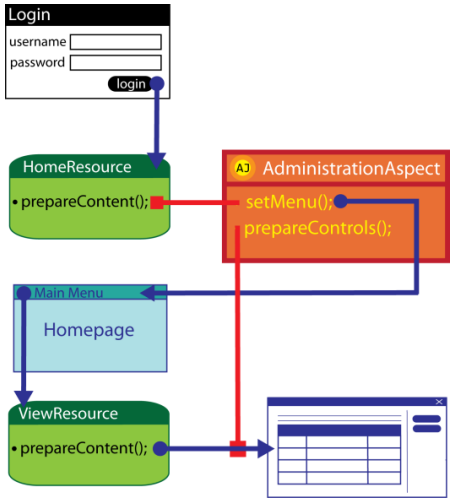


Figure 6. Aspects can change the content before pages are rendered

Listing 1 shows an aspect that modifies the main menu before it is presented to the user (one possible scenario for this is filtering main menu items based on user permissions created by Kroki administration subsystem). Since generic web engine is designed as one-page Ajax-based web application, once the user is logged in, the interaction in it's entirety takes place on the home page and Restlet resource in charge of this page. So, as mentioned before, in order to modify the main menu preparation process, we need to attach our aspect to `prepareContent` method of `HomeResource` class. Freemarker template looks up main menu list by the name `main_menu`, so it will be the name by which we will put our modified menu into `dataModel`. Listing 1 represent basic steps described above.

VI. CONCLUSIONS

The paper presented architectural solutions incorporated in the Kroki tool which enable prototype execution. Core elements of this rapid prototyping technique are the application repository and generic engines which create functional enterprise Java application based on the developed specification.

```
public aspect MainMenuAspect {
/* Create the pointcut that
   intercepts PrepareContent method in
   Home resource and obtain home resource object
*/
  public pointcut setMenu
        (HomeResource homeResource) :
  call(public void HomeResource.prepareContent())
        && this(homeResource);


  after (HomeResource homeResource) :
        setMenu(homeResource) {
        //Obtain main menu list from AppCache
        ArrayList<AdaptMenu> menus =
        AppCache.getInstance().getMenuList();

        //Do something...

        //Put modified main menu to data model
        homeResource.addToDataModel
                ("main_menu", menus);
}
```

Listing 1. Aspect extension example

The process does not include traditional code generation techniques where complete programming code (or most of it) is being generated. In our approach, engines are generic enough to cover basic business operations on provided data.

The engines based on aspect-oriented programming enable: (1) easier inclusion of future tools and features that can affect the specified enterprise application execution; (2) integration of generated and hand-written code during application evolution, when its code is exported to a general-purpose programming tool; (3) dynamic adaptation of application's look and behavior in accordance with the user rights defined in Kroki's administration subsystem.

Decision to use this approach is guided by years of research in model-driven development and generic enterprise systems [1, 13, 14, 15], which resulted in development of our EUIS DSL [13] and Kroki tool.

Now, we are planning to convey a research as large as possible including end-users, business specialists, students, and IT experts in order to get their feedback and measure their reactions while using the tool. We plan to base our experiment on Methodology Evaluation Model (MEM), which represents software engineering specific extension of earlier Technology Acceptance Model (TAM). The MEM model suggests that a certain methodology, in order to be successfully accepted and used, needs to satisfy both subjective and objective measures of usefulness and ease of use. In the experiment, we plan to measure performance based variables (actual efficiency and actual effectiveness) using cognitive load measurement approach, and perception based variables (perceived ease of use and perceived usefulness), like presented in [16].

REFERENCES

[1] G. Milosavljevic, M. Filipovic, V. Marsenic, D. Pejakovic, I. Dejanovic, Kroki: A mockup-based tool for participatory development of business applications.. SoMeT (p./pp. 235-242), : IEEE. ISBN: 978-1-4799-0419-8, 2013.

[2] T. Cerny, M. Macik, M.J. Donahoo, J. Janousek, Efficient description and cache performance in Aspect-Oriented user interface design, Computer Science and Information Systems (FedCSIS), 2014

[3] T. Cerny, K. Cemus, M. J. Donahoo, and E. Song. Aspect-driven, Data-reflective and context-aware user interfaces design. Applied Computing Review, 13(4):53–65, 2013

[4] J. L. H. Agustin, P. C. del Barco, A model-driven approach to develop high performance web applications, Journal of Systems and SoftwareVolume 86, Issue 12, 2013

[5] J. M. Rivero, J. Grigera, G. Rossi, E. Robles Luna, F. Montero, M. Gaedke. 2014. Mockup-Driven Development: Providing agile support for Model-Driven Web Engineering. Inf. Softw. Technol. 56, 6, June 2014

[6] J. Solano, Exploring How Model Oriented Programming can be Extended to the UI Level, PhD Thesis, University of Ottawa, 2010, http://hdl.handle.net/10393/28569

[7] A. Forward, O. Badreddin, T. Lethbridge, J. Solano, Model-driven rapid prototyping with Umple, Software: Practice and Experience, Volume 42, Issue 7, pages 781–797, July 2012

[8] H. Störrle, Model-driven development of user interface prototypes: an integrated approach, Proceedings of the Fourth European Conference on Software Architecture: Companion Volume, pp 261-268, Copenhagen, Denmark

[9] T. Buchmann, Towards Tool Support For Agile Modeling: Sketching Equals Modeling, *Proceedings of the 2012 Extreme Modeling Workshop*, pp. 9-14. ACM, 2012.

[10] A. Coyette, S. Schimke, J. Vanderdonckt, C. Vielhauer, Trainable Sketch Recognizer for Graphical User Interface Design , *Human-Computer Interaction–INTERACT 2007*. Springer Berlin Heidelberg, 2007. 124-135

[11] A. Coyette, J. Vanderdonckt, In *Human-Computer Interaction-INTERACT 2005*, pp. 550-564. Springer Berlin Heidelberg, 2005.

[12] B. Plimmer, M. Apperley, Interacting with Sketched Interface Designs: An Evaluation Study, In *CHI'04 extended abstracts on Human factors in computing systems* (pp. 1337-1340). ACM.

[13] B. Perišić, G. Milosavljević, I. Dejanović, B. Milosavljević, "UML Profile for Specifying User Interfaces of Business Applications", Computer Science and Information Systems,Vol. 8, No. 2, pp. 405-426., 2011

[14] G. Milosavljević, B. Perišić, "A Method and a tool for rapid prototyping of large-scale business information systems", Computer Science And Information Systems, Vol. 02, pp. 57-82, 2004

[15] B Milosavljevic, M. Vidakovic, S.Komazec, G. Milosavljevic, User interface code generation for EJB-based data models using intermediate form representations,2nd International Symposium on Principles and Practice of Programming in Java, PPPJ 2003, Kilkenny City, Ireland, 2003

[16] S. Abrahão, E. Insfran, J. A. Carsí, M. Genero, "Evaluating requirements modeling methods based on user perceptions: A family of experiments", *Information Sciences*, Volume 181, Issue 16, pp. 3356-3378 (2011)

[17] Kroki, www.kroki-mde.net

[18] Kroki demo, http://youtu.be/r2eQrl11bzA

Software development with Scrum – Telenor Serbia E-Business Success Story

Marčelja Aleksandar*, Makitan Vesna**, Ivković Miodrag **
*Telenor d.o.o, Belgrade, Republic of Serbia
marcelja@gmail.com
**University of Novi Sad/Technical Faculty "Mihajlo Pupin", Zrenjanin, Republic of Serbia
vesna@tfzr.uns.ac.rs, misa.ivkovic@gmail.com

***Abstract*— In order to achieve modern business requests such as shorten time of advancement and launching the new products and services as well as end users inclusion in product development process, specific methodologies need to be applied.**

This paper presents application of the Scrum methodology in software project realization. It is the real world Telenor Serbia E-business software project described from every aspect of the Scrum implementation beginning with the project start, then through pilot project, sprint execution, usability testing and concluding remarks. This may be used as an example in other businesses and companies in order to improve project realization and business operations.

I. INTRODUCTION

General project management concept has proven itself through numerous successful projects and in that way, may be applied in project realization in any area. Thus, PMBOK (Project Management Body of Knowledge) standard, as project management methodology basis, may be applied in any area. Concerning specificities of certain projects, like in software development projects, many companies use other methodologies such as: PRojects IN Controlled Environments (PRINCE2), agile methodologies, Rational Unified Process (RUP) framework or Six Sigma methodologies. [1]

This paper concerns agile methodologies, and especially the Scrum methodology. The fact that during the last decade of the past century information technology (IT) projects had success rate of 16.2% [2], was taken into account. According to [2], this success percentage increased to 35% till the 2006. This may be attributed to the developed and applied methodologies in IT projects realizations.

Agile methodologies present adaptive software development, which means that projects are mission driven, based on components and use time cycles in order to achieve preset deadlines. [3, 4, 5, 6, 7] Furthermore, all the agile methodologies (Scrum, Extreme Programming (XP), Feature Driven Development (FDD), Lean Software Development, Agile Unified Process (AUP), Crystal, and Dynamic Sistem Development Method (DSDM)) have iterative flows and increment software deliverables in short iterations. [1]

The main goal in this paper was to report a successful use of Scrum in a company that has not previously used it and, as a result, an improvement in its business operations. The Scrum methodology was described through the real Scrum roles, Scrum flow and Scrum artifacts (product backlog and sprint backlog) of the E-business project realized in the Telenor Serbia company. This is described in the following chapters from: How it all started, through Pilot project, Sprint execution, Demo day and Usability testing to the Conclusion.

Described project does not represent only one of the kind, with determined duration, but the Scrum as accepted way of work for all e-business initiatives, i.e. requests. It is an ongoing project. Conclusion remarks show how many features have been realized, and how many sprints have been done since the beginning of the implementation. Used tools contain information classified as "confidential" and may not be externally shared.

II. RELATED WORK

In the modern world, USA mostly, agile methodologies are greatly used. One of the biggest indicators about this is number of business ads where experts from this area are mentioned [8]. The data about agile methodologies implementation in Serbia dates from 2010. They are mostly available at portals of IT companies [9], banks [10], news [8, 11], and agencies for trainings [12, 13] or expert conferences [14] about agile methodologies. There is a research [15] about agile methodologies analysis in software companies that confirms Scrum implementation in Serbia. However, there are no sources, or they are unavailable, about Scrum methodology application experiences in Serbia. This emphasized importance of this paper, which by described experience may positively influence on bigger Scrum methodology application in our country, as well as on increasing success rate and company competitiveness, which uses this methodology.

III. HOW IT ALL STARTED?

There was a need for faster software products delivery and, on the other hand, the need for starting implementation immediately. Considering this in 2014, E-business team of Telenor Serbia decided to implement agile way of work in order to execute online initiatives. The pace of advancement and launching the new services and products as well as including the end users in product development process, are essentially important for the success in the field of e-business. In that way extraordinary users experience and high quality product would be provided at the same time. Previously used project management concept included traditional methodologies with phased approach (waterfall model) have not provided satisfying results. This concept implies

that all the requirements are predefined and that end user gives inputs at the barely start and then during the test phase, after more than a few weeks or months of analysis and implementation, which is not enough user involvement. These reasons influenced on the decision to apply totally different approach.

E-business team wanted to launch the new self-care portal for Telenor mobile network users in short time, but in the moment of project initiation only the part of the user requirements was known. Online market demands that the development from idea to the final product lasts only a few weeks. Considering the fact that the e-business market is extremely dynamic and things may be changed on the daily basis, it was necessary to enable continuous development, which will provide totally new requirements realization as well as, changes during the implementation process if it is needed, without delays in ongoing projects or initiating the new ones. It was necessary to implement requirements as they appear, based on market change or needs of other stakeholders. There were a lot of requirements, between twenty or thirty per month, but such requirements never seemed important enough to start the special project realization. That is why those requirements have never been realized, or had a realization with enormous delay. In order to avoid that delay in requirements realization where implementation lasts a few weeks or even months, traditional project management and planning process had to be abandoned at all costs. Telenor favors end users and wants to put people and user experience in the first place, and extraordinary user experience could only be delivered through constant and fast end users feedback and usability testing during the implementation process. This would enable the most effective way of high quality products development that at the same time totally satisfies end users demands. E-business team considered its needs and plans with the project portfolio office as well as with the software development department. Limited number of internal resources and large number of ongoing projects at the company's level imply that the traditional approach does not enable self-care portal realization in timely and economic satisfying manner. It was clear and without cost analysis that buying the on the shelf commercial solution does not pay off and this option was not at the table at all. Giving up of the new portal realization was not considered too, and the new solution needed to be found. Decision was made to try with the agile methodology application.

IV. THE PILOT PROJECT

In order to continue the work it was necessary to get approval from the company's top management. The challenge was even bigger because in that moment, company's knowledge about agile approach in initiatives execution was only theoretical, without any previous experience (except one software developer that was certified ScrumMaster). Due to time constraint there were two parallel fronts: gaining top management approval and learning by doing (learning about the methodology, researching, training, etc.). ScrumMasters organized internal workshops for sharing knowledge between colleagues. Two ScrumMasters and one Product Owner were certified additionally, and there were a lot of help from the other business units inside the Telenor Group that had pre experience in agile approach. Gaining the knowledge about the methodology lasted for two weeks, and at the same time top management gave their consent for the start of six months pilot project. After six months evaluation of success rate will be done as well as the decision about further steps and continuing of application.

Product Owner role was assigned to the manager of the E-business team (the main business stakeholder). The greatest support to the Product Owner during the selection of features, that should be developed were many marketing experts from different e-business areas. One certified programmer took the role of the ScrumMaster. Considering assignments at the other ongoing projects, company did not have enough resources for software development, and it was decided that during the pilot project company engage outsource company for the development services. Instead of engaging market research agency, it was decided to save money and costs and gain better results through immediate and real time feedback and usability testing with future users during the development process.

V. THE FIRST SPRINT

The objective was to launch the new portal in beta version as soon as possible, to realize only the set of features necessary for launching, and the rest of features as well as the future requests should be realized on the fly. Two weeks after the initial idea it was time for the first sprint. In order to arrange sprint planning together with the business stakeholders Product Owner prepared Product backlog (the list of required features).

ScrumMaster is not obligatory participant in Product backlog creation, but he/she may assist and be consulted if it is needed. Workshop for Product backlog update usually lasts one working day and then every business stakeholder adds his/her requirements to the list. Taking into account importance and order of realization they set priority for every requirement. Requirement in the Product backlog explains the need considering business approach, and it has:

- feature name,
- system that needs to be changed (where it should be),
- feature description and interaction between the system and the end user (user story),
- detailed description with the business rules that should be implemented and guidelines for the user interface design (description),
- example of the drawing or the print screen, which is optionally, and
- priority that is needed.

Based on the completed Product backlog the entire Scrum team, i.e. Product Owner with his/her team, ScrumMaster and development team perform grooming of the Product backlog. In the start they eliminate requirements whose realization is not possible, or the ones not enough well defined, in order to avoid losing time on its reconsideration during the sprint planning. Grooming lasts for a few hours, or one working day at most. The next step is the one-day workshop for sprint planning that includes the entire Scrum team, as well as the business stakeholders, owners of the features from the Product backlog.

The sprint duration is previously defined and usually lasts up to two or three working weeks. It starts with requirements analysis from the beginning according to the list with requirements sorted by priority, starting with the ones with the highest priority. Every stakeholder, i.e. requirement owner represents his/her requirement (user story, description, etc.). At the same time user story represents success criteria that is used, after the sprint execution, as success verification of requirement realization.

When business stakeholder represents his/her requirement, ScrumMaster explains, i.e. translates this requirement to "technical language" for the development team in order to represent work scope, technology that will be used (programming language, platform, etc.) and what they should do exactly. In case that feature realization includes development that is not in responsibility of the development team (for example, development team develops portal, and feature demands integration with backend system that needs change too) Product Owner is responsible for resources that are needed from the Company and who will execute this development. When development team members confirm that they understand feature completely, the duration assessment for the feature realization starts. Duration assessment means that every member of development team raises card with number that symbolizes estimation of days needed for him or her to realize each individual feature. Every development team member gives his/her personal assessment without consulting other team members. If there are no deviations between individual assessments, the assessment that is agreed among development team members, which is realistic for the specific feature, is taken into account (average value of given assessments or the highest one).

If there are great differences, team members with the extreme assessments explain their evaluation. After this, assessment that has the most valuable explanation is taken into account. Duration assessment is done for the entire Product backlog. In case that some of the features are impossible to realize, or they are not enough clear, too complex or have correlation with the activities that are not in the Scrum team authority and it will not be completed in time, they get lower priority level, and its realization is possible later. In case that there is a feature that cannot be realized even during the entire sprint, it decomposes on simpler requirements. After duration assessment needed for realization of every feature, team creates sprint backlog based on number of the development team members, time needed for the feature realization and previously defined number of days for the sprint execution. Bonus i.e. backup features are chosen too. Development team needs to know what to do in case that some feature completes earlier, or if its realization is for some reason cancelled. The output of the sprint planning workshop is Sprint backlog including tools for tracking of realization (burndown chart) and it is regularly updated by the ScrumMaster during the sprint execution. Later, during the sprint execution, team members are choosing the tasks and making agreements about assignment of concrete tasks.

VI. SPRINT EXECUTION

Every day of sprint execution is the same and starts with the daily standup meeting. It is a short meeting that gathers Product Owner, ScrumMaster and development team. The meeting lasts ten minutes at most and development team members inform Scrum team about the work that is done the day before, activity plan for that day and, if it is needed, they discuss about any problem they have in realization in order to make decision and take steps for overcoming the problem. If some of the development team members have a problem that cannot be solved by Scrum team itself (for example, enable access to the testing environment), ScrumMaster takes over responsibility for solving the problem so the team can continue with work.

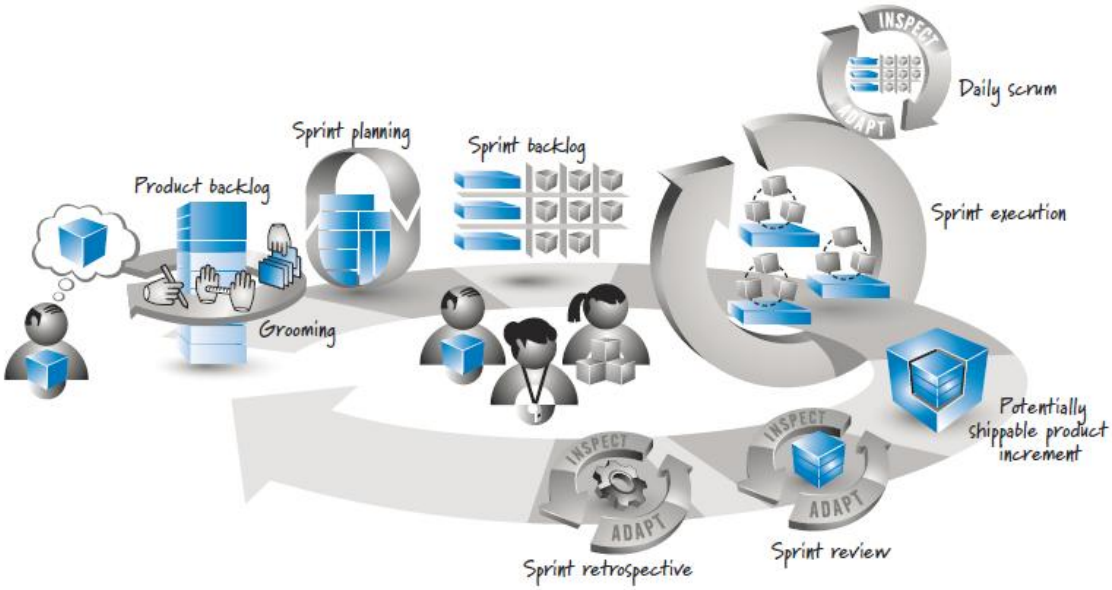


Figure 1. Scrum activities and artefacts

Product Owner may use his/her influence to help problem overcoming as soon as possible. In the beginning, Microsoft Excel spreadsheet was used for progress tracking purposes, but in the meantime it was replaced by open source tool named *Trello* that appeared as more useful and easier for updating. Also, this tool enables interactivity and communication between Scrum team members in situations when a team member needs additional information about requirement from the business stakeholder. The member of the development team may use *Trello* on daily basis and send question to the business stakeholder without development delays till the next day standup meeting. If development team member start feature realization that lasts for three days, and during its realization requirement change appears, demanded by business stakeholder, then this stakeholder and development team member may consult each other if it is possible to implement changed feature in the previously set timeline. If this realization is not possible, then at the next day standup meeting Product Owner and Scrum team together are making decision about extending the deadline for realization of this feature. Consequence of this decision may be to cancel realization the last requirement (requirement with the lowest priority) in the list from this sprint. Also, business stakeholder has possibility to decide about:

- canceling realization of that feature in the ongoing sprint – this is in case when its realization without change does not have meaning;
- continuing realization of the next requirement in the list, or if it is acceptable
- continuing feature realization as it was planned in the first place.

Change request in the middle of the sprint is not well and it should be avoided at all costs in order to maintain pace of the development team. After the feature realization, member of development team marks this feature as completed and goes on with the next requirement according to the plan. Then, business stakeholder checks developed feature and, in case of noticing the bug he/she marks that in *Trello*. By doing that business stakeholder notifies development team member that worked on the feature development to make necessary corrections. In that way, during the development and in real time, testing and corrections are made as well. This decreases the risk of having any bugs that could jeopardize success of the entire sprint to the lowest possible level. Sprint execution has pre-defined duration and ends on the planned date, no matter if all features are realized in time or not. If during the sprint execution all the features are realized before deadline, development team takes predefined bonus i.e. backup goals and starts their realization. During the sprint execution Product owner and his/her team work on creating and gathering new requirements that will be candidates for realization in the future sprint. Figure one shows the most common Scrum activities and artefacts.

VII. DEMO DAY

After the last day of sprint execution Demo day is organized. It is an event that gathers all members of Scrum team, as well as business stakeholders. Demo day duration depends on time needed for demonstration of realized features. Scrum master and development team members demonstrate realized solution; show final product that is actually working software that runs in testing environment. At the same time verification is done too. It has to be checked if the feature is realized in accordance with the User story, other criteria agreed during the sprint planning and changes during the sprint execution. Product owner confirms if the feature is realized as is should be, i.e. that the requirement is successfully completed. If all planned requirements are realized successfully, without bonus requirements, sprint is declared as successful. If even one planned feature did not realized successfully sprint is unsuccessful and those features transfer to the next sprint. Feature deployment may be realized right after sprint execution if it was a small improvement or a patch that is not connected to other activities (for example, launching product at other channels, marketing campaign, etc.), but its launching may be planned for the later specific date. Right after the sprint execution there is a new cycle, starting with grooming, sprint planning, etc.

VIII. USABILITY TESTING

After the feature development and deployment there is a usability testing conducted by Usability expert who works with end users. Usability testing goal is to identify possibilities for feature improvements based on the way that users use features and feedback from the end users. If it is necessary to make certain correction or feature improvement, it would be defined as requirement for some of the following sprints.

IX. CONCLUSION

During two weeks only, E-business team succeeded that from bringing idea to making decision, gain enough basic knowledge, necessary approvals from the top management and to start with agile approach application. The first sprint started at the beginning of the second quarter in 2014, and its goal was to launch the new selfcare portal as soon as possible. Beta version of the portal was launched for less than two months, after three successful sprints. Pilot project was declared as successful after six months, because it is satisfied three previously defined success criteria:

1. Deliverables on platform in the scope;
2. Use findings to create guidance for other similar initiatives, and for creation of development handbook;
3. Explore how agile methodology functions in Telenor.

The fact that Scrum is convenient for software project development mostly is used in this project realization. This means that there is no parallelism, which enables simultaneous work on many different projects. For example, if company has one senior expert for business support development systems, who works at four or five ongoing projects at the same time with 10-20% of availability, he/she cannot be member of the Scrum team. It is not possible to accomplish portfolio project roadmap with projects that follows waterfall methodology. To overcome this organizational challenge it was needed to engage external developers.

After described project Scrum become default way of work for all projects in the e-business area. In the beginning of 2015 there is fourteenth sprint in the role, and by late January over 170 new features were realized at Telenor selfcare portal, web shop, smartphone application,

etc. Without agile methodology application and Scrum approach the most of these features would never be realized. Now, there is a continuous and effective development with team members who are completely committed to the e-business initiatives, unlike before, when experts from different areas were assigned to too many different projects, which disrupted their focus and efficiency. Development team is externally engaged and availability of internal resources that depends on other projects practically does not influence the plan and activities of the E-business team. For the first time there is a fast and direct business communication, i.e. marketing service with programmers at the daily basis, and in that way there are no delays or errors due to bad and untimely communication. New features are launched constantly, and involvement of end users directly into the development process enabled launching the high quality solutions, that brings extraordinary user experience to even higher level.

ACKNOWLEDGMENT

This research is financially supported by Ministry of Education and Science of the Republic of Serbia under the project number TR32044 “The development of software tools for business process analysis and improvement”, 2011-2015.

REFERENCES

[1] Kathy Schwalbe, "Information technology project management" sixth edition, 2008, Course Technology, Boston, USA

[2] www.standishgroup.com

[3] Ken Schwaber, "Agile Project Management with Scrum", 2004, Microsoft Press

[4] Kenneth S. Rubin, "Essential Scrum: A Practical Guide to the Most Popular Agile Process", 2012, Addison-Wesley

[5] Lyssa Adkins, "Coaching Agile Teams: A Companion for ScrumMasters, Agile Coaches, and Project Managers in Transition", 2010, Addison-Wesley

[6] Ken Schwaber & Jeff Sutherland, "Scrum Guide", 2013, Scrum.org

[7] Robert K. Wysocki, "Effective Project Management", seventh edition, 2014, John Wiley & Sons, Inc.

[8] “Tržište rada počinje da vrednuje agilne kvalifikacije“, http://www.agilni.rs/index.php?limitstart=5, 27.1.2015.

[9] “Novosti, jun, 2010”, http://www.asp.rs/, 27.1.2015.

[10] “Meni najbliža”, http://mobilnimarketing.me/domace-aplikacije/meni-najbliza-mobilna-aplikacija-komercijalne-banke/, 27.1.2015.

[11] “Agilna populacija u Srbiji koristi Scrum metodologiju”, http://www.naslovi.net/2014-12-03/ekapija/agilna-populacija-u-srbiji-koristi-scrum-metodologiju/12541096, 27.1.2015.

[12] “Novo: Scrum kursevi u Eccentrix-u”, http://www.eccentrix.rs/novosti/novo-scrum-kursevi-u-eccentrix-u, 27.1.2015.

[13] “Scrum sertifikacija”, https://sr-rs.facebook.com/pages/Puzzle-Software/111788045550235, 27.1.2015.

[14] “Sinergija 14: Implementacija Scrum-a na Balkanu”, http://forum.benchmark.rs/showthread.php?339231-Sinergija-14-Implementacija-Scrum-a-na-Balkanu, 27.1.2015.

[15] “Analiza primene agilnih metodologija u softverskim organizacijama vojvođanskog IT klastera”, Medaković, Đ. Departman za poslovnu informatiku, Ekonomski fakultet Subotica, http://www.slideshare.net/PositiveNoviSad/analiza-primene-agilnih-metodologija, 27.1.2015.

Developing distributed multi-core and many-core architecture using java agents

Jelena Tekić, Predrag Tekić, Miloš Racković
University of Novi Sad, Faculty of Sciences, Novi Sad, Serbia
{radjenovic, tekic} @uns.ac.rs, rackovic@dmi.uns.ac.rs

***Abstract*—In this paper java agent architecture for utilizing available multi-core and many-core hardware is presented. Architecture is developed using JADE (Java Agent DEvelopment Framework) and OpenCL standard for programming multi-core and many-core devices. A JADE-based system can be distributed across available machines and OpenCL promises portability of the developed code between heterogeneous devices. OpenCL is an open, royalty free, standard developed by Khronos group for parallel programming of heterogeneous devices (CPU's, GPU's, …) from different vendors. Developed java agents communicate with each other in accordance with FIPA specification and are independent from operating system of the specific machine they run on. In this paper two java agents are presented. First agent inspects available hardware with OpenCL support and sends message to the second agent which runs simulation utilizing all the available devices with OpenCL support discovered by the first software agent.**

Keywords: JADE, OpenCL, multi-core, Java Agent, GPU

I. INTRODUCTION

Recently, there has been a breakthrough in computer processor technology. Multi-core and many-core processors are replacing single core processors. Graphics Processing Units (GPUs) have an important role in today's high performance computing applications. High performance computing was a privilege of small group of scientists with budget to fund expensive large computer clusters or specially manufactured High performance computers. Nowadays, with massively produced multi-core and many-core processors it is available to almost every commodity desktop/personal computer.

Large processing power potential of new multi-core and many-core processors (CPUs and GPUs) has attracted researchers and developers to start considering to utilize power of those new processors in solving their specific scientific problems.

There are number of scientific fields which are profiting from this trend in computer hardware industry, among them is computational fluid dynamics (CFD). This field of science in recent years, is starting to make use of new and more powerful multi-core and many-core processors in order to solve more and more complex numerical simulations.

In this paper we have proposed software solution which will be able to automatically detect, and make use of any device available for large scientific calculations and therefore reduce time needed for specific scientific calculation. Proposed solution should create, at runtime, a cluster of hardware devices able to perform parallel numerical computations and execute specific task on all of the available devices in that cluster.

In the following pages we will explain software technology we have chosen for our software solution and the architecture of the proposed software solution.

A. FIPA

The Foundation for Intelligent Physical Agents (FIPA) is an international organization helping to promote the industry of intelligent agents. FIPA is developing specifications supporting interoperability among agents and agent-based applications.

FIPA Abstract Architecture includes:

- Distributed computing platforms or programming languages,
- Messaging platforms,
- Security services,
- Directory services, and,
- Intermittent connectivity technologies.

Also FIPA Abstract Architecture defines the following concepts:

- A model of services and discovery of services available to agents and other services,
- Message transport interoperability,
- Supporting various forms of ACL representations,
- Supporting various forms of content language, and,
- Supporting multiple directory services representations.

FIPA abstract architecture specification describes how to make the agent system, agents and services that rely on the system, in Figure 1. Mapping FIPA-abstract architecture on the concrete realization is shown.

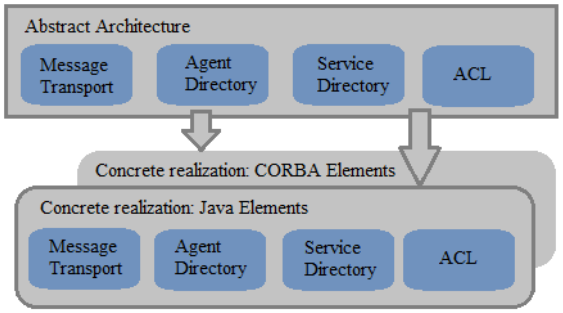


Figure 1. Mapping FIPA-abstract architecture on the concrete realization

Agents living in one environment can communicate by sending messages to each other. Messages are encoded in an Agent-Communication-Language.

Services provide support services for agents. Services are generally implemented as software that is accessed via method invocation. Service root is provided to an agent on start-up, service root will provide a set of service-locators such as: message-transport-services, agent-directory-services and service-directory-services.

Agent-directory-service provides a location where agents register their descriptions as agent-directory-entries. Agent-directory-entry consists of at least two key-value-pairs Agent-name and Agent-locator. Agent-name is globally unique name and Agent-locator consists of one or more transport-descriptions. Each transport-descriptions contain transport-type, transport-specific-address and zero or more transport-specific-properties. Also agent-directory-entry may contain other descriptive attributes (services offered by the agent, cost of using the agent, restrictions on using the agent, etc.)

Service-directory-service provides discovering of services and a location for registration of service-directory-entries. Service-directory-entry consisting of at least one key-value-pairs: Service-name, Service-type and Service-locator. Service-locator contains one or more key-value tuples that consists of a signature type, service signature and service address. Also service-directory-entry may contain other descriptive attributes (cost of using the service, restrictions on using the service, etc.)

Agent message consists of sender name, receiver names and message content. Every message must have one sender and zero or more receivers. Message content can be determined by the ontologies, in Figure 2 message structure is shown.

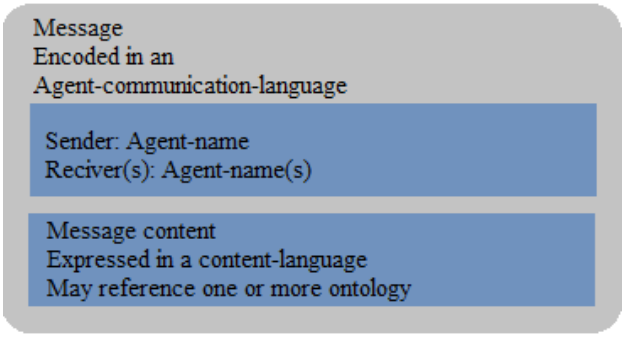


Figure 2. Message structure

Before transport message is encoded into a payload and then included in a transport-message. Transport-message consists of payload and envelope. Envelope defines transport-description: sender address, receiver address, transport protocol, encoding-representation etc.

In order to obtain the parameters needed to transport messages agent communicate with directory agent. The parameters that are needed are placed in the agent locator. Communication begins by a search of the directory service for the agents (1 : Query) . Based on the name of the destination agent, the starting agent from the directory agents obtains agent-directory-entry which contains all the necessary parameters for communication (Transport - type, Transport- specific -address , Transport- properties) . The initial agent can send a message by the first type of transport (e.g. , HTTP protocol) , and then change the type of transport (e.g. SMTP protocol) if the destination agent is able to communicate using both.

FIPA specification supports two types of security mechanisms: message validation and encryption of the messages. Validation of the message includes the ability to detect changes of the messages that occurred after sending the message. If some of the protection mechanisms are applied in the description of the message (envelope) additional parameters must be placed. These additional parameters will allow the use of protective mechanisms. [1,2]

B. JADE

JADE (Java Agent DEvelopment Framework) is free software Framework that is complied with the FIPA specifications and implemented in the Java language.

JADE was initially developed by the University of Parma. Copyright holder and distributor for JADE software is Telecom Italia.

JADE consists of run-time environment, a library of classes (used by programmers for developing agents) and graphical tools for administrating and monitoring the activity of running agents.

Container is running instance of the JADE runtime environment, it can contain several agents. Platform contains several containers; one container in platform is special container called Main container. Main container is always active in a platform, all other containers register with it; it contains special agents AMS (Agent Management System) and DF (Directory Facilitator). The AMS (Agent Management System) provides the naming service and represents the authority in the platform. DF (Directory Facilitator) provides service for finding agents.

JADE provides communication between agents by use of asynchronous message passing, The JADE asynchronous message passing paradigm is shown in Figure 3. For each agent JADE runtime posts messages sent by other agents in the agent message queue and notify agent that he got message. When the agent will process message is completely up to the programmer. [3,4]

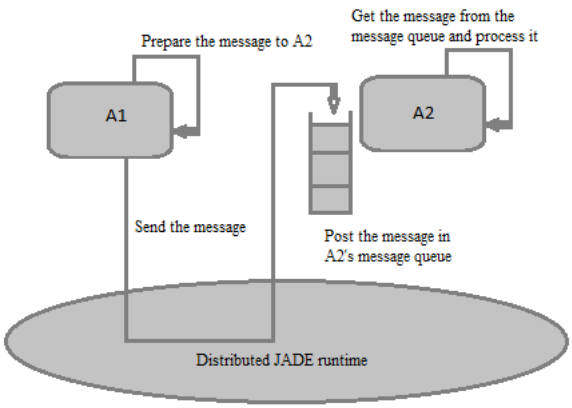


Figure 3. The JADE asynchronous message passing paradigm

C. OpenCL and JOCL

Open Computing Language (OpenCL) [2] is an open, royalty free, standard developed by Khronos group [3] for parallel programming of heterogeneous devices (CPUs, GPUs, DSPs) from different vendors. The execution model for OpenCL consists of the controlling host program and kernels which execute on OpenCL devices.[8]

JOCL is library that provides Java binding for OpenCl. JOCL is very similar to original OpenCL API. JOCL functions are written as static methods. Semantics and signatures of JOCL methods are consistent with the original OpenCL library functions, except for the language-specific limitations of Java. [5]

II. IMPLEMENTATION DETAILS

In this chapter we will focus on describing implementation of java software agents which are used to inspect hardware devices and start numerical simulation. Benchmark problem that is used for the numerical simulation is well known and often used in CFD, lid driven cavity flow.[6,7] Implementation details of this benchmark simulation problem will not be subject of this paper.

In order to start numerical simulation on the available hardware devices, first we need to inspect available hardware resources in the specific environment. Java software agent *InspectPlatformAgent* has been created for that purpose.

InspectPlatformAgent examines all available devices on the specific platform and verifies which devices have support for OpenCL specification and print that information to the console. Software agent creates message, which will be sent to another agent, with information about devices which have support for the OpenCL specification (no device – *NO*, all devices – *ALL*, only GPU – *GPU*, only CPU - *CPU*). In order to create agent in JADE agent framework, agent needs to inherent *jade.core.Agent* class and implement *setup()* method. Code that is used to examine available devices implemented as a part of *setup()* method. Implemented code uses JOCL library and OpenCL API functions to inspect hardware devices. *InspectPlatformAgent* has inner class which provides functionality of sending messages to another agent. Created class listing is shown in Figure 4. This class creates new *behaviour*, *TickerBehaviour* have been used, which periodically (in this example every 10 seconds) repeats task that need to be carried out.

```
TickerBehaviour  loop = new TickerBehaviour(this,10000) {
    public void onTick() {
        ACLMessage msg = new ACLMessage(ACLMessage.INFORM);
        msg.addReceiver(new AID ("simulationExample", AID.ISLOCALNAME));
        msg.setContent(mesContent);
        myAgent.send(msg);
        ACLMessage msgRecive = myAgent.receive();
        if (msgRecive != null) {
         doDelete();
        }
    }
} ;
addBehaviour(loop);
```

Figure 4. InspectPlatformAgent class listing

InspectPlatformAgent have a task to send ACL message. Message is created by setting message type (ACLMessage.INFORM in this case) local name of the agent that receives message and content of the message.

Another java software agent have been created *SimulationExampleAgent* to carry out numerical simulation on the discovered devices. This agent receives previously created message about platform from *InspectPlatformAgent.* After message have been processed it sends notification to *InspectPlatformAgent* which triggers *doDelete()* command which will destroy *InspectPlatformAgent* software agent.

SimulationExampleAgent software agent receives message from the first agent, and based on that message starts numerical simulation on all of the available hardware devices with OpenCL support and prints out time needed to carry out simulation on each device.

A. Java agent environment

In order to start software agent we need to compile it with following commands:

javac -classpath lib\jade.jar;lib\JOCL-0.1.9.jar -d classes src\examples\inspectPlatformAgent\ SimulationExampleAgent.java

and

javac -classpath lib\jade.jar;lib\JOCL-0.1.9.jar -d classes src\examples\inspectPlatformAgent\ InspectPlatformAgent.java

Following command:

java -cp lib\jade.jar;lib\JOCL-0.1.9.jar;classes jade.Boot –gui

is used to start JADE user interface with dependencies need in this example (*lib\JOCL-0.1.9.jar*). In the Figure 5 JADE user interface is shown.

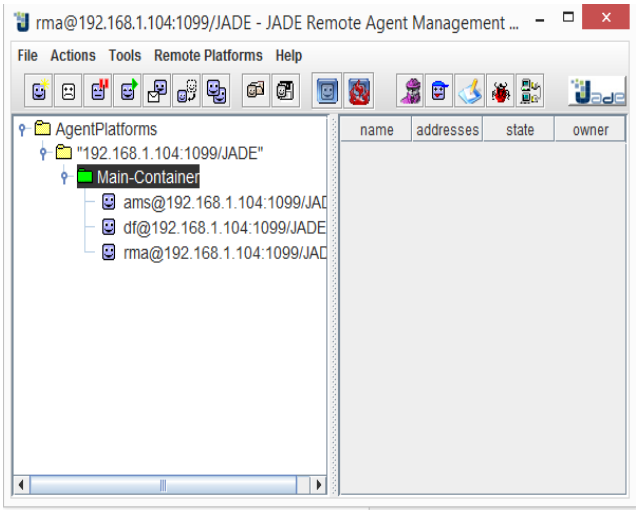


Figure 5. JADE user interface

Upon selecting main container, software agents can be started. In the list of the available (compiled) agents, previously created and compiled container *examples.exampleAgent.InspectPlatformAgent* is selected as shown on Figure 6.

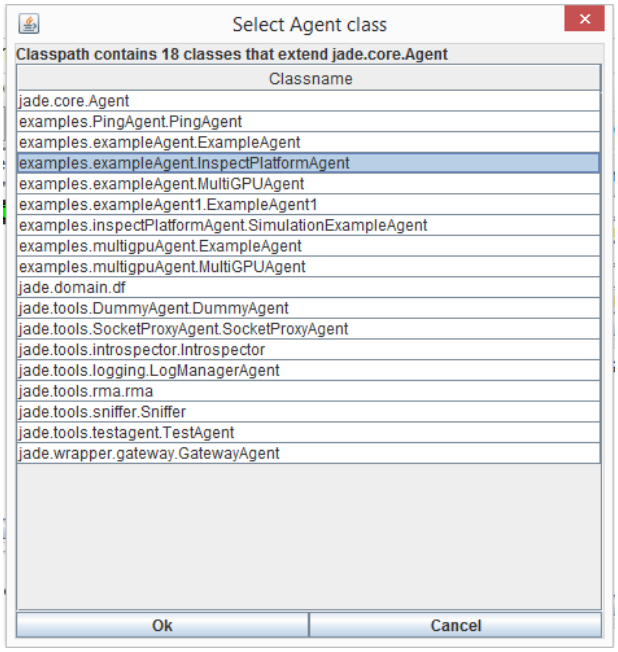


Figure 6. List of the available (compiled) agents

In the next step, each agent needs to be started separately. List of all compiled software agents will appear in user interface, as shown in Figure 7. Result of started agents is console output shown in Figure 8.

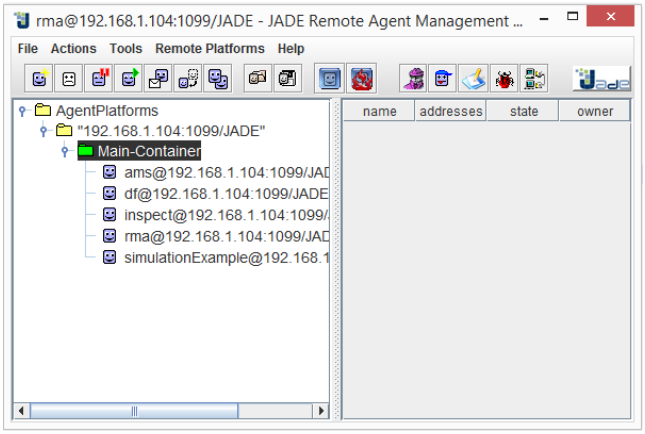


Figure 7. JADE user interface with started agents

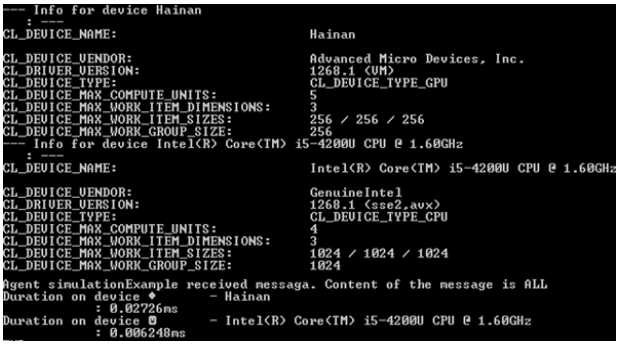


Figure 8. Result of started agents

III. CONCLUSION

In this paper we have described software architecture created to utilize all the available hardware devices with OpenCL support, in order to carry out parallel numerical simulation. Software architecture is based on java agents, using JADE agent framework and java OpenCL library (JOCL) for programming multi-core and many-core devices with support for OpenCL specification. Created software agents are used to inspect hardware devices and after having that information to run numerical simulation on all the available devices, utilizing computational power of all discovered devices.

It has been shown that it is possible to develop software architecture that is both platform and vendor independent which can be used to discover and use all devices with parallel computing potential. This solution can be used to carry out scientific computations by using all the available devices in specific computer network.

ACKNOWLEDGMENT

The work is partially supported by Ministry of Education and Science of the Republic of Serbia, through Project OI174023: ‘‘Intelligent techniques and their integration into wide-spectrum decision support’’

REFERENCES

[1] FIPA http://www.fipa.org/, January 2015.

[2] FIPA specification http://www.fipa.org/specs/fipa00001/SC00001L.pdf, January 2015.

[3] JADE http://jade.tilab.com, January 2015.

[4] JADE Programming Tutorilal http://jade.tilab.com/doc/tutorials/JADEProgramming-Tutorial-for-beginners.pdf, January 2015.

[5] Java OpenCL Library - JOCL, http://www.jocl.org/, January 2014.

[6] Tekić, P., Rađenović, J., Lukić, N., Popović, S., Lattice Boltzmann simulation of two-sided lid-driven flow in a staggered cavity, International Journal of Computational Fluid Dynamics, (ISSN 1061-8562), Vol. 24, Issue 9, pp. 383-390, 2010.

[7] Tekić, P., Rađenović, J., Racković, M., Implementation of the Lattice Boltzmann Method on Heterogeneous Hardware and Platforms using OpenCL, Advances in Electrical and Computer Engineering, (ISSN: 1582-7445), Vol. 12,No 1, pp. 51-56, 2012.

[8] Rađenović, J., Tekić, P., Racković, M., VISUALISATION OF FLOW AND TEMPERATURE FIELD CALCULATED BY LB METHOD IN POST-PROCESSING SOFTWARE PARAVIEW, Proceedings of the International Conference on Information Society Technology and Management (ICIST), pp. 303-306,2014. ISBN: 978-86-85525-14-8

Semantic search framework for distributed semantically based cheminformatics and bioinformatics datasets

Branko Arsić*, Marija Đokić*, Vladimir Cvjetković*, Petar Spalević**, Siniša Ilić**
* Faculty of Science, Kragujevac, Serbia
** Faculty of Technical Sciences, Kosovska Mitrovica, Serbia
brankoarsic@kg.ac.rs, m.djokic@kg.ac.rs, vladimir@kg.ac.rs, petar.spalevic@pr.ac.rs, sinisa.ilic@pr.ac.rs

***Abstract*— Integration of dissimilar and heterogeneous data sources is a continuous challenge for life science investigation. For the researchers in a domain of life science, finding the relevant information across different data sources is of crucial importance. Current researches strongly depend on availability, accessibility and effective data use. Semantic Web technologies enable facilitated collected data aggregating. Extracting data from large, semantically based datasets became a trying challenge for a researcher who needs to put an effort in discovering each connection in the datasets of interest, different properties among datasets and classes, name conventions, as well as to understand precise meaning of every enumerated part. For systems without specific conventions in organizing separate datasets, these tasks can be quite complex and time-consuming. In this paper, we present the challenges of data integration of same-domain datasets and the use cases that identify our approach aimed at finding the relevant data essential for our further research. We have developed a semantically based web application that utilizes the ontologies and SPARQL queries for the data process search and integration in molecular biology research center. For selected substance from our semantic dataset we have generated SPARQL queries for data discovering by means of the existing and constructed templates. In this way a knowledge database for valid and tested SPARQL queries is created and presented in a new ontology, established for this purpose.**

I. INTRODUCTION

Different institutions present their data in different manners, using diverse nomenclature and structure presentation. The data with equivalent meaning are presented in various formats and storages. A significant amount of knowledge is not available to all biomedical researchers, even though they have a frequent need to use datasets derived from other distributed systems. Most traditional integration methods are not scalable and based on static mapping approaches which results in significant access to information processing constraints. The most serious issue, caused by the constant expansion of novel data sources, is analyzing of disconnected data, which has become a threat for future successful and purposeful explorations. Such a quick-growing environment asks for updates monitoring worldwide. Bearing in mind all the things previously written, it is evident that the institutions with similar goals are faced with challenges when finding and comparing published data due to the fact that it is often necessary to interpret large amounts of data.

The current integration systems possess a lot of shortcomings such as inconsistent terminology, various data formats and customary alterations in data models. These shortcomings are transferred to the domain of biology and chemistry as well. Certain shortcomings in the domain of interest are dealt with in the section 3 with emphasis on difference between separated datasets. Semantic Web technologies have mechanisms to reduce the burden of data integration and sharing [1]. Semantic Web offers to its users well-defined models for data aggregation of heterogeneous data sources using explicit semantics among notions, with clear interlinks and relations - all of these packed with simplified annotation and with the possibility of publishing knowledge on web. With precise notion meaning, we receive an opportunity to connect data in different locations, unrelated at first glance, but sharing the biological and chemical domains. Ontologies are created as mechanisms for connecting similar data sources, offering a possibility to search heterogeneous datasets through a single SPARQL query [2][3]. Combination of ontologies and federated SPARQL queries for data retrieval can significantly simplify modeling of arbitrary concept and data structures and implementation of required functions. Numerous organizations use semantic web technologies to build ontologies as assistance in data integration and search processes. For instance, large initiatives (EBI, DrugBank, OpenPHACTS, PubChem,…) present their data by means of Semantic Web context, since they constantly updating the data structure.

Real system that is supported in this paper is the Research Center (RC) for testing of active substances [4]. The Center is also the forerunner of a large Project financed by the Ministry, and the subject of its analysis includes monitoring of in vitro effects of active substances in the cell lines of different origin (primarily cancer cell lines) and primary cells isolated from different tissues [5]. Active substances that are tested in laboratories are candidates for medicaments prior to being approved for medical treatments. Tests include measuring of the effectiveness of a substance in inhibiting a specific biological function (IC50) in human cancer cell lines, the mechanisms of apoptosis, migration and angiogenesis. The results of work in Center are experiments with complex structure that present complex relationships among various terms and concepts from the

Center work area. The structure is expected to further expand in the future, so it requires flexible modeling and representation that can be easily updated. These observations have led us to the semantic web technologies as appropriate choice. References [6][7] present earlier developed PIBAS (Preclinical Investigation of Bioactive Substances) ontology for data storing of active substances used in complex experiments, model systems, cancer cell-lines and experimental results.

Nowadays data sources are being developed in individual form isolated one from another. This leads to heterogeneous and challenging compute environment. Process of finding information about entity of interest in distributed systems can be difficult and laborious. Researchers are forced to follow the path between poorly connected sources. To achieve this aim, data about cell-lines, targets, proteins, pathways, drugs, and chemical compounds (substances) must be efficiently integrated and accessible to all the researchers. If a researcher wants to create valid SPARQL queries, he will have to discover almost all possible interlinks between notions. For many researches, this task can be a real challenge. Process of discovering new drugs, binding targets with cell lines, connecting pathways with active substances or compounds, relating genes with different diseases and similar processes can be very difficult.

Real need for application upgrading comes from the fact that some laboratories can test the same substance, but with different cancer-lines in different conditions. This complementary data can be very useful for QSAR analysis and very good direction for the future experiments. Some initiatives can be focused in other experiments and in other parameters such as pathways and targets. Obtained information is precious, because the search process has tendency to save researchers' time and resources. It will be shown later in examples how to get various experimental results for the same substance. In order to help life science community in finding relevant information we developed web application based on Semantic Web technologies. This application enables searching process within cheminformatics and bioinformatics datasets, based on federated SPARQL queries. Consequently, the application allows easy extraction of relevant information from all available, distributed sources. In the same time, this presents one kind of integration between datasets of interest. General endpoint is created, so in one place we have a possibility to explore some large systems in specific manner (see section 4). Everybody can create the templates and make them available for research community.

This paper is organized in the following way: The second section gives an overview of the literature and software in the domain of data integration in biology and chemistry. The third section talks about challenges in data integration and motivation for this work. The fourth section describes software's architecture as data integration mechanism between local ontologies and distributed systems that use templates and developed ontology. One part of this section is dedicated to federated SPARQL queries as a main product of software. Conclusion contains short survey of paper key points and directions for future work.

II. Related Work

Semantic Web technologies have potential to bridge before mentioned difficulties because they offer a common framework which permits data to be shared and reused across different systems. These technologies have a promising role in the field because they enable data integration and interlinking. Every initiative dealing with biology and chemical data has developed tools for data visualization and extracting on its own. In the following paragraph we present some of these large integrated systems and tools from our domain. Wild et al. indicated the importance of data integration in cheminformatics and bioinformatics [8]. Examples of successful exploitation and integration using Semantic Web technologies in biology can be seen in papers [9][10]. Ontologies as a main part of Semantic Web are finding their way in many areas of life sciences, especially in biomedicine [11].

Various initiatives for data integration of chemical and biological sources using a Semantic Web context have been established in the past decade. Open PHACTS represents Semantic web approach for addressing bottlenecks to drug discovery, developed as a shared platform for integration [12]. The European Bioinformatics Institute (EBI) provides freely available data from life science experiments, performs basic research in computational biology [13]. EBI developed Java based web application LODEStar as a generic SPARQL endpoint and Linked Data browser to provide a consistent interface and some enhanced functionality for querying and browsing EBI based dataset. Among many other services and tools, EBI offers UniChem API [14] as free available service which allows mappings of small molecule based on adopted and stable standard, InChIs and InChIKeys. Chem2Bio2RDF [15] is a solution based on Semantic web technologies which covers around 25 different datasets related to chemical/biology needs. This solution includes data about genes, compounds, drugs, pathways, side effects, diseases, and MEDLINE/PubMed documents. SLAP [16] tool for drug prediction is made by the same initiative. An aim of OpenTox community is to develop an interoperable predictive toxicology framework which may be used as an enabling platform for the creation of predictive toxicology applications [17]. For these purposes ToxPredict (http://www.toxpredict.org/) and ToxCreate (http://www.toxcreate.org/) applications are developed. Multiple chemical-protein annotation resources integrated with diseases and clinical outcomes information are presented by ChemProt integration system [18]. In recent years, various biological and chemical initiatives resulted in many tools and frameworks, including ChESS [19], WENDI [20], Data Dryad [21], IsaTab [22] and OpenTox.

Among large systems there are many identical datasets included. With this application we have direct access to all datasets which is positive since we can never be sure if a dataset is up to date.

III. Challenges in Integration of Biological and Chemical Data

Refined knowledge discovery process in life science includes moving through large numbers of interlinks among variant data sources. Researchers need to access several databases to perform tasks and identify potentially

constructive data, following each change manually, which is a time-consuming and error-prone process.

Current level of data integration is often facing syntactic and semantic heterogeneity challenges. The appearance of the same dataset in more than one global integration systems with different name convention, for the same compound, is probably the most common problem. For example, compound ZINC00120249 in Chem2Bio2RDF is denoted as 65-22-5 in ChemSpider integration model. With this notation we cannot be sure, without checking, if these substances are same or not.

In many situations compounds found in Chem2Bio2RDF and in our dataset are not available in some other dataset. This is caused by the fact that datasets within large system are not up to date or they have not treated substance yet. We need to be careful in creating queries because the results can be empty sets and thus get a wrong impression. Absence of substance in a result of a query is a good sign that tested substance is new on the market and that we are on the right track perhaps.

Problem in the forming of SPARQL queries is often followed by the absence of adequate endpoints. Many of them can be unsuitable, temporally unavailable and cause the lack of adequate results. In that case the problem is solved by using valid, alternative endpoints.

Large world initiatives take a few base datasets at the beginning and create the one-way relations to other internal datasets saving resources. Different integrated systems can have connected datasets but with different relation direction. In almost all situations one correct query is not correct for other systems. At the same time, different names and properties are used for the same notions.

IV. System Architecture

Our approach represents a collection of the templates from different initiatives from a domain of interest which are connected within developed core ontology. This approach enables an easier and faster way for searching included data sources than existing applications which are focused on a system which they belong to, rather than encompassing several initiatives in the same time. Every semantically based initiative invested a lot of time for presenting their work through the different SPARQL queries. Here, we want to avoid exploring datasets by using existing and self-created SPARQL queries. By the term "template" we shall mean every take over, valid SPARQL query. We gathered and connected them into unique storage in the form of ontology. New templates formed during the period of exploring these datasets are added as well.

Architecture of our web application (see Fig. 1) can be divided into three logical parts (layers): 1. User-interface for selecting input parameters and target integrated systems. Next step is filtering and selecting possible templates for selected systems 2. SPARQL framework represents an engine for adjusting templates into one federated query 3. Query execution on FUSEKI server and data retrieving. In the following paragraphs these layers are elaborated.

In Fig. 2 we can see the application interface. As possible input, molecular formula and weight are general parameters and the results offer the data of several substances. Sometimes this can cause confusion and make it difficult to track specific information. There are also two notations of chemical structures: *SMILES* and *InChI/InChIKey* strings. SMILES strings are omnipresent in chemical world, but they do not cope with every needed entity specification. Also, they are frequently not unique and cause the relevant data to misplace during the search. Conversely, the InChI and InChIKey (hashed InChI) strings lately gained a leading role among unique identifiers. Number and type of input parameters depend on the existing template's options and can be extended in the future.

As mentioned earlier, semantically based web application is developed with pillars in form of ontology (see Fig. 3). Developed ontology consists of information such as URI, endpoint and tested patterns, for every integrated system. Presented semantically based integrated systems have extendable list of possible templates for extracting ontology structures. For every template three parts pattern is constructed: a) *description* – used for user interface as a description of what we obtain as a result by embedding corresponding query b) allowed input parameter c) *query (template)* – SPARQL code with empty input parameter value. With input parameter choice and selected target system, web application filters possible corresponding templates that can be used in SPARQL generator layer. There is an option to choose several input parameters and integrated systems at the same time. Some of the integrated systems which are in focus in this paper are Chem2Bio2RDF, Bio2RDF, CHEBI, ChEMBL, DrugBank, LODD and

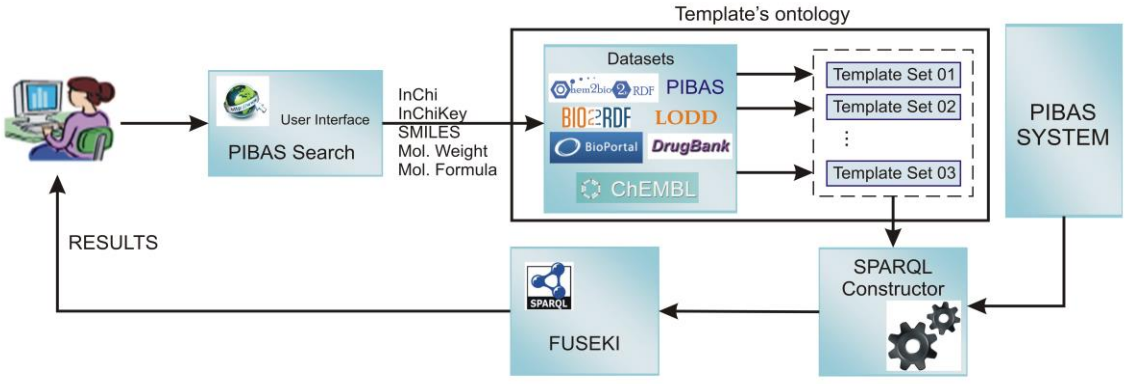


Figure 1. Web application arhitecture. Selected parameters determine final SPARQL query.

BioPortal. Possible templates are limited programmatically and a user can select a template according to given description.

Selected first layer arguments are forwarded to SPARQL constructor which generates federated SPARQL query with our input parameters' value. An example of such queries can be seen in Fig. 4. If we select more systems, we will obtain result with data from several datasets. Some of the data are complementary, and some are redundant. Different systems can have different standards for the same notion and we cannot say with certainty, in constructor runtime, that one notion is equal to another. After query is generated, an application sends it to FUSEKI server with general endpoint for execution.

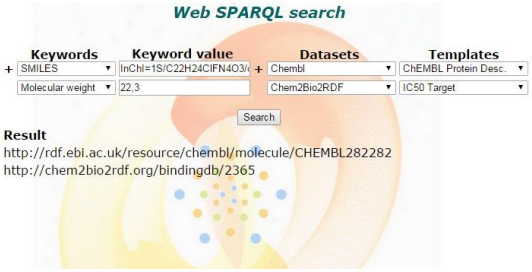


Figure 2. Application interface

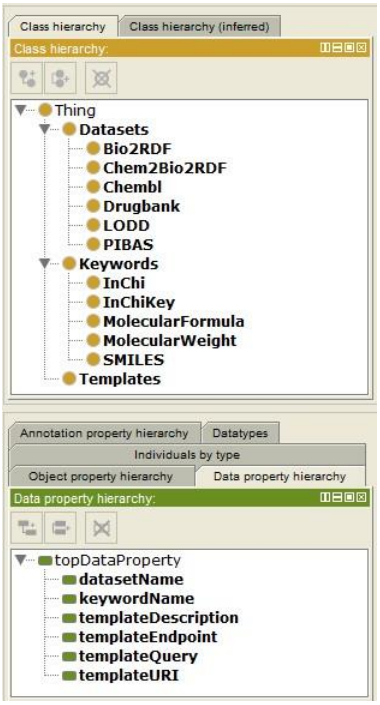


Figure 3. Ontology for integrated system's templates

V. CONCLUSION

The main activities of the Research Center include chemical synthesis, purification and extraction of bioactive substances, microbiological, cell and molecular, immunological and pharmacological preclinical testing of active substances. It has been proven, and in practice confirmed, that certain classes of chemical structures show larger or smaller biological effect. Data stated above can indicate whether the special attention should be paid to the new substance or not, thus avoiding time and resource consumption. The existence of such a database and of the corresponding software implies that the communication in the opposite direction exists as well, where CPCTAS could suggest to chemists a new synthesis direction. On one hand, in this way we can gain extra knowledge about substances through different kinds of experiments, in various conditions. On the other, we can have a new untested compound which could be important in terms of scientific research. The confirmation for this assumption stems from an empty result set. Fast accumulation of new, not-up-to-date data and reorganization of existing data can cause major problems in knowledge discovering. With new software we can facilitate the research work. Our application can be a standard for other small laboratories which want to improve the work and save resources.

For the future work we plan to deal with redundant data of the application results. For example, substance with molecular formula $C_{20}H_{12}N_2O_2$ and Chembl synonym 2'-Phenyl-[2,4']Bibenzooxazolyl is denoted as *CHEMBL113081*. The same substance within Chem2Bio2RDF initiative is denoted as *m180094*. However, this problem could be overcome with InChI/ InChIKey value. Different standards for the same notion are apparent in the case of data/object properties naming. Substance's target property within Bio2RDF initiative is *gene-name* from DrugBank namespace, and target property within Chem2Bio2RDF initiative is *geneSymbol* from Uniprot namespace. This problem is much bigger and requires deeper exploration of all the large systems and laborious mapping processes. On the other side, every initiative deals with different experiments focusing on specific results. Complementary data are also important, and with redundancy presents "all in one pack" problem with similar steps in their solving.

```
PREFIX pibas:<http://cpctas-lcmb.pmf.kg.ac.rs/2012/3/PIBAS#/>
PREFIX compound: <http://chem2bio2rdf.org/pubchem/resource/>
PREFIX pubchem: <http://chem2bio2rdf.org/pubchem/resource/>
PREFIX chembl: <http://chem2bio2rdf.org/chembl/resource/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?compound ?cell_line_name
FROM <http://chem2bio2rdf.org/pubchem>
FROM <http://chem2bio2rdf.org/chembl>

WHERE{
?compound compound:std_inchi ?compound_inchi.
?compound_inchi pibas:InChi ?inchi.
?compound_inchi pibas:MolecularWeight ?mol_weight.

SERVICE<cheminfov.informatics.indiana.edu:8890/sparql>
{
?activities chembl:molregno ?compound;
            chembl:standard_value ?standard_value;
            chembl:standard_units ?standard_units;
            chembl:assay_id ?assay_id.

?assay2target chembl:assay_id ?assay_id.
?assay2target chembl:tid ?cell_line.

?cell_line chembl:pref_name ?cell_line_name.
}
FILTER regex(?cell_line_name,"cancer","i").
FILTER(?mol_weight="223.5").
FILTER(?inchi="InChI=1S/C22H24ClFN4O3/c1-29-20-13-19-16(12-21(20)31-8-2-5-28-6-9-30-10-7-28)22(26-14-25-19)
27-15-3-4-18(24)17(23)11-15/h3-4,11-14H,2,5-10H2,1H3,(H,25,26,27)").
}
```

Figure 4. Generated SPARQL query

ACKNOWLEDGMENT

This paper was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (scientific projects TR32023, III41010, ON174033 and III44006).

REFERENCES

[1] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web", Scientific American, vol. 284 (5), pp. 29-37, May 2001.

[2] N. Guarino, D. Oberle, and S. Staab, "What is an Ontology?", In *Handbook on Ontologies*, Springer Berlin Heidelberg, 2009.

[3] J. Pérez, M. Arenas, and C. Gutierrez, "Semantics and Complexity of SPARQL," In *The Semantic Web-ISWC 2006*, Springer Berlin Heidelberg, pp. 30-43, 2006.

[4] CPCTAS-LCMB, Faculty of Science, University of Kragujevac, Serbia, http://cpctas-lcmb.pmf.kg.ac.rs.

[5] Project data http://cpctas-lcmb.pmf.kg.ac.rs/lcmb/pibasEn.htm.

[6] V. Cvjetković, M. Đokić, B. Arsić, and M. Ćurčić, "The ontology supported intelligent system for experiment search in the scientific research center", *Kragujevac Journal of Science*, vol. 36, pp. 95-110, 2014.

[7] B. Arsić, M. Đokić, V. Cvjetković, P. Spalević, M. Živanović, and M. Mladenović, "Integration of bioactive substances data for preclinical testing with Cheminformatics and Bioinformatics resources," *Proceedings of the 23nd International Electrotechnical and Computer Science Conference (ERK 2014)*. pp. 146-149.

[8] D. J. Wild et al., "Systems chemical biology and the Semantic Web: what they mean for the future of drug discovery research," *Drug discovery today*, vol. 17(9), pp. 469-474, 2012.

[9] H. Min et al., "Integration of prostate cancer clinical data using an ontology," *Journal of biomedical informatics*, vol. 42(6), pp. 1035-1045, 2009.

[10] D. Salvi et al., "Merging Person-Specific Bio-Markers for Predicting Oral Cancer Recurrence Through an Ontology," *Biomedical Engineering*, IEEE Transactions on, vol. 60(1), pp. 216-220, 2013.

[11] B. Smith et al., "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration," *Nature biotechnology*, vol. *25*(11), pp. 1251-1255, 2007.

[12] A. J. Williams et al., "Open PHACTS: semantic interoperability for drug discovery," *Drug discovery today*, vol. 17(21), pp. 1188-1198, 2012.

[13] S. Jupp et al., "The EBI RDF platform: linked open data for the life sciences," *Bioinformatics*, vol. *30*(9), pp. 1338-1339, 2014.

[14] J. Chambers et al., "UniChem: a unified chemical structure cross-referencing and identifier tracking system," *Journal of Cheminformatics*, vol. 5(3), 2013.

[15] B. Chen et al., "Chem2Bio2RDF: a semantic framework for linking and data mining chemogenomic and systems chemical biology data," BMC bioinformatics, vol. 11(1), 255, 2010.

[16] B. Chen, Y. Ding, and D. J. Wild, "Assessing drug target association using semantic linked data," *PLoS computational biology*, vol. 8(7), e1002574, 2012.

[17] B. Hardy et al., "Collaborative development of predictive toxicology applications," *J. Cheminform*, vol. 2(1), 2010.

[18] O. Taboureau et al., "ChemProt: a disease chemical biology database," *Nucleic acids research*, vol. 39, pp. 367-372, 2011.

[19] L. L. Chepelev and M. Dumontier, "Chemical entity semantic specification: knowledge representation for efficient semantic cheminformatics and facile data integration," *J. Cheminform.* vol. 3(20), 2011.

[20] Q. Zhu et al., "WENDI: a tool for finding non-obvious relationships between compounds and biological properties, genes, diseases and scholarly publications," *J. Cheminform.* vol. 2(6), 2010.

[21] J. Greenberg, "Theoretical considerations of lifecycle modeling: an analysis of the dryad repository demonstrating automatic metadata propagation," *Inheritance, and value system adoption. Catalog. Classif. Quart.* vol. 47, pp. 380–402, 2009.

[22] S. A. Sansone et al., "Toward interoperable bioscience data," *Nat. Genet.* vol. 44, pp. 121–126, 2012.