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

FICT

FACULTY OF INFORMATION AND
COMMUNICATION TECHNOLOGIES
"ST. KLIMENT OHRIDSKI"
UNIVERSITY – BITOLA,
MACEDONIA

AIIT

International Conference

PROCEEDINGS OF THE 7TH INTERNATIONAL CONFERENCE ON APPLIED INTERNET AND INFORMATION TECHNOLOGIES

WWW.AIITCONFERENCE.ORG

ZRENJANIN, SERBIA, OCTOBER 5-6, 2017

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

Faculty of Information and
Communication Technologies
University of St. Clement Ohridski
Bitola, MACEDONIA
http://fikt.edu.mk/

International Conference on Applied Internet and Information Technologies ICAIIT 2017

P R O C E E D I N G S

Zrenjanin
October 5-6, 2017

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

Faculty of Information and Communication Technologies
University of St. Clement Ohridski
Bitola, Macedonia

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

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

Technical preparation and design:
Vladimir Brtka, Dejan Lacmanović, Predrag Pecev, Željko Stojanov

Cover design:
Predrag Pecev

CIP - Каталогизација у публикацији
Библиотека Матице српске, Нови Сад

004(082)

INTERNATIONAL Conference on Applied Internet and Information Technologies (7 ; 2017 ; Zrenjanin)
Proceedings [Elektronski izvor] / [7th] International Conference on Applied Internet and Information Technologies ICAIIT 2017, Zrenjanin, October 5-6, 2017 ; [organizer] Technical Faculty "Mihajlo Pupin", Zrenjanin. - Zrenjanin : Tehnički fakultet "Mihajlo Pupin", 2017. - 1 elektronski optički disk (CD-ROM) : tekst ; 12 cm

Nasl. sa naslovnog ekrana. - Str. [VI]: Introduction / Vladimir Brtka.
- Bibliografija uz svaki rad.

ISBN 978-86-7672-304-1
1. Technical faculty "Mihajlo Pupin" (Zrenjanin)
a) Информационе технологије – Зборници

COBISS.SR-ID 317671687

.

International Scientific Committee

Mirjana Pejic Bach, University of Zagreb, Croatia
Androklis Mavridis, Aristotel University of Thessaloniki, Greece
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
Zeljko Jungic, ETF, University of Banja Luka, Bosnia and Hercegovina
Saso Tamazic, Univerisity of Ljubljana, Slovenia
Marijana Brtka, Centro de Matemática, Computação e Cognição, Universidade Federal do ABC, São Paulo, Brazil
Zoran Cosic, 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, Belgrade, Serbia
Zlatko Čović, Subotica Tech / Department of Informatics, Subotica, Serbia
Diana Gligorijevic, 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 - president
Branko Markoski, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Dalibor Dobrilović, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Željko Stojanov, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Ilija Hristoski, Faculty of Economics - Prilep, Macedonia
Dejan Lacmanović, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Zoltan Kazi, 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

Jelena Stojanov, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia

Violeta Manevska, Faculty of Information and Communication Technologies - Bitola, Macedonia

Aleksandar Markoski, Faculty of Information and Communication Technologies - Bitola, Macedonia

Igor Nedelkovski, Faculty of Information and Communication Technologies - Bitola, Macedonia

Ilija Jolevski, Faculty of Information and Communication Technologies - Bitola, Macedonia

Blagoj Ristevski, Faculty of Information and Communication Technologies - Bitola, Macedonia

Natasa Blazeska-Tabakovska, Faculty of Information and Communication Technologies - Bitola, Macedonia

Elena Vlahu-Gjorgievska, Faculty of Information and Communication Technologies - Bitola, Macedonia

Zoran Kotevski, Faculty of Information and Communication Technologies - Bitola, Macedonia

Tome Dimovski, Faculty of Information and Communication Technologies - Bitola, Macedonia

Mimoza Bogdanoska-Jovanovska, Faculty of Information and Communication Technologies - Bitola, Macedonia

Andrijana Bocevska, Faculty of Information and Communication Technologies - Bitola, Macedonia

Snezana Savoska, Faculty of Information and Communication Technologies - Bitola, Macedonia

Nikola Rendevski, Faculty of Information and Communication Technologies - Bitola, Macedonia

Bozidar Milenkovski, Faculty of Information and Communication Technologies - Bitola, Macedonia

Pece Mitrevski, Faculty of Information and Communication Technologies - Bitola, Macedonia

Željko Eremić, College of Technical Sciences - Zrenjanin, Serbia

Organizing Committee

Vladimir Brtka, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia - president
Pece Mitrevski, Faculty of Information and Communication Technologies - Bitola, Macedonia
Miodrag Ivković, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Biljana Radulović, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Ivana Berković, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Branko Markoski, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Željko Stojanov, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Dalibor Dobrilović, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Eleonora Brtka, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Dejan Lacmanović, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Ljubica Kazi, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Višnja Ognjenović, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Zoran Kotevski, Faculty of Information and Communication Technolo-gies - Bitola, Macedonia
Ilija Jolevski, Faculty of Information and Communication Technologies - Bitola, Macedonia
Blagoj Ristevski, Faculty of Information and Communication Technolo-gies - Bitola, Macedonia
Natasa Blazeska-Tabakovska, Faculty of Information and Communica-tion Technologies - Bitola, Macedonia
Violeta Manevska, Faculty of Information and Communication Technol-ogies - Bitola, Macedonia
Tome Dimovski, Faculty of Information and Communication Technolo-gies - Bitola, Macedonia
Nikola Rendevski, Faculty of Information and Communication Technol-ogies - Bitola, Macedonia
Zoltan Kazi, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Predrag Pecev, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
Nebojša Tatomirov, Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia

INTRODUCTION

New Internet and Information Technologies create new perspectives in many domains, while the way we collect and use information has dramatically changed. The focus is on facilitating implementation of the Internet and Information Technologies in all areas of human activities. This conference is a great opportunity for researchers, as well as for users in many areas, to share their experiences and to disseminate their knowledge. The conference is organized as cooperative effort of our Technical faculty "*Mihajlo Pupin*" from Zrenjanin, Serbia and Faculty of Information and Communication Technologies "*St. Kliment Ohridski*" from Bitola, Macedonia. Through years, many new ideas, solutions, research results were published on this conference, mainly in the domain of IT. Authors and coauthors are coming from many foreign countries and Serbia. These are: Hungary, Macedonia, Bosnia and Herzegovina, Canada, United States, Romania, Russia, Bulgaria, Montenegro, Germany and India. We thank them all for their contribution to the conference.

The objectives of the **International Conference on Applied Internet and Information Technologies** are aligned with the goal of regional economic development. 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 theoretical, technical and methodological point of view, and deals with the analysis, the design, the applications and realization of various technologies.

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

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

President of the Organizing Committee
Ph.D Vladimir Brtka

Zrenjanin, October 2017

ORGANIZATOR WITH PARTNERS:

MAIN ORGANIZATORS

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

Faculty of Information and Communication Technologies
University of St. Clement Ohridski
Bitola, MACEDONIA
http://fikt.edu.mk/

ORGANIZATION PARTNERS:

ИРКУТСКИЙ
НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ
ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Faculty of Computer Science
Irkutsk National Research Technical University
Irkutsk, RUSSIA
http://www.istu.edu/structure/57/9518/1801/

INTERNATIONAL CONFERENCE
ON APPLIED INTERNET AND INFORMATION TECHNOLOGIES
OCTOBER 5-6, 2017

PAPER INDEX

Robust solutions for complex systems

Péter Odry[1], Mihály Klincsik[2], Zoltán Sári[2], Zoltán Góbor[3], Ervin Burkus[1], István Kecskés[1], Zoltán Király[1], Ákos Odry[1], Vladimir Tadić[1], Zoltán Vízvári[2]

[1]Dunaújváros University, Department of Information technology
2401 Dunaújváros, Táncsics Mihály u 1a, Hungary
{podry, kecskes, burkus, odrya, tadity, kiru}@uniduna.hu
[2] Pécs University, Polák Mihály Department, 7624. Pécs, Hungary
{vizvariz, klincsikm, ski.lod}@gmail.com
[3] Bavarian State Research Center for Agriculture (LfL),
85354 Freising, Vöttinger Str. 36, Germany
gobor.zoltan@o2mail.de

Abstract. This material presents a comparative analysis about two systems that seemingly cannot be compared. These systems are the hexapod walking robot and tomography procedures. Their common properties can be defined the following ways: high uncertainty while creating the model, there are a lot of disturbing phenomena. So-called "inverse" solutions should be used in order to get essential descriptive information from the system. We need to use robust solutions to achieve the aforementioned goals. The robustness property will be presented through a hexapod walking robot example.

Keywords: robust solutions, inverse problem, hexapod robot, tomography.

1. Introduction

For a longer time we have been researching and developing in the following fields:
- Hexapod robot walker (theoretical mechanics) [1] and [2].
- Tomography (theoretical electromagnetics) [4] and [5].

In both cases (see Fig. 1), we are solving inverse tasks in a, particularly nonlinear environment.

In the field of theoretical electromagnetics, we are working on tomography research, in essence in the area of medical-biological electric impedance tomography (we have a patent in this topic – [5]). We are also doing magnetic impedance tomography researches which specifically means testing the internal structure of „high-thickness" ferromagnetic materials (we have a patent – [4]). For example, in the case of depth analysis of nuclear reactor walls or for any kind of other, industrial-like uses where other types of nondestructive analyzes cannot be applied. In both of these cases of tomography research, we used our own high sensitivity electronics/sensor systems and self-developed calculating processes.

In the field of theoretical electromagnetics, we are working on tomography research; in essence in the area of medical-biological electric impedance tomography (we have a

patent in this topic – [5]). We are also doing magnetic impedance tomography researches which specifically means testing the internal structure of „high-thickness" ferromagnetic materials (we have a patent – [4]). For example, in the case of depth analysis of nuclear reactor walls or for any kind of other, industrial-like uses where other types of nondestructive analyzes cannot be applied. In both of these cases of tomography research, we used our own high sensitivity electronics/sensor systems and self-developed calculating processes.

Fig. 1. A) Magnetic impedance tomography for inspection of ferromagnetic materials [4]; B) Hexapod walker robot Szabad(ka) II [1]

There is a high degree of uncertainty in both research topics, but the reasons are different. In case of tomography researches, while doing measurement processes we can encounter systematic errors, which most of the time are irreducible. The noise, that occurs while measuring, is the biggest problem in further processing. While solving the electromagnetic inverse task, not only the measurement interference signals cause issues but the "mathematical" noise (quantization error) also adds up to the uncertainty. Dimensional "depth" information content should be created in case of tomography, this is the main problem. We can only make measurements on the surface of the body; we cannot do inner measurements, because this is a nondestructive process. In the field of theoretical electromagnetics, direct task means that we know the examined body's material structure. While using Maxwell equation and exciting the body, we can calculate the examined body's surface's voltage conditions. This is also a complex task (nowadays, there are finite element FEM packs, for example, COMSOL). Compared to this, inverse task solving is a lot more complex. This is because there are only finite measuring results on the surface and this is how we need to determine the material's structure in depth.

In case of Hexapod walking robot, direct task means that we have to completely describe the simulations, which includes every nonlinear, nonstationary etc. phenomenon. With the help of this simulation, every property can be calculated. When we have 18 joints (with 18 drives) it is also not an easy task. The "inverse" task means the following: we can measure a finite number of points; here we need to mention that we are not able to measure every single, important parameter (these properties can be

about structural, like mechanical properties). We need to emphasize that this is a real structure and the system is especially nonlinear and nonstationary.

The basis of searching for the robust solution is that the optimal solutions need to be found. Not every optimal solution is a robust solution, it can even be different. On the multi-dimensional Pareto surface, the solutions exclude each other by their behavior on other surfaces. This means that we should be prudent when choosing the optimal solution because it needs to be robust as well.

Within our article we are going to represent robustness on a very simple example in case of the Hexapod walking robot: on the simplified robot leg system. Fuzzy is usually independent of any kind of use which requires controlling technique task solutions. There can be another case, when next to the not properly chosen [3] parameters, due to its basic philosophy, we get the robust solution as a result, not like other methods (in this article we do not get into proving this fact). This is why in the following chapters we only deal with the devices which are mentioned in the chapter.

One of our basic expectations was to have a reliably working system while doing the system's research and realizing process, in hardware and software environment as well. We wanted it to function under any scenario. In theory, the number of scenarios is infinite if the system is for free use. In case of an endless number of scenarios, the test cannot be done. This means we need to create a set of simple scenarios that mostly cover opportunities arising in real situations on their own. Or if they have any other realizable combination, then they are combined. The main goal is to have robustness in the system's wider area, let it be foreseeable or unforeseeable environmental effects. However, it is difficult to realize this in most cases.

2. Theoretical Background

We are going to do a really short introduction about the applied theoretical background; due to the limited scope we are not able to discuss it in depth. First, we are going to deal with the tomography research, after that we are going to deal with the robot.

2.1. EIT Tomography Basics

This non-local property of conductivity imaging, which still applies at the moderate frequencies used in EIT, is one of the principal reasons that EIT is difficult. It means that to find the conductivity image one must solve a system of simultaneous equations relating every voxel to every measurement. Non-locality in itself is not such a big problem provided we attempt to recover a modest number of unknown conductivity parameters from a modest number of measurements. Worse than that is the ill-posed nature of the problem. Small errors in measurement can violate consistency conditions, such as reciprocity [4]. The tarting point for consideration of EIT should be Maxwell's equations. But for simplicity let us assume direct current or sufficiently low a frequency current that the magnetic field can be neglected.

The time harmonic Maxwell's equations:

$$\begin{aligned} \nabla \times \vec{E} &= -i\omega\mu\vec{H} \\ \nabla \times \vec{H} &= i\omega\varepsilon\vec{E} + \vec{J} \end{aligned} \quad . \tag{1}$$

With combining the conductivity and permittivity as a complex admittivity $\sigma + i\omega\varepsilon$, the solution is in next form

$$E = -\nabla\phi \, . \tag{2}$$

In the mathematical literatures we will often see the assumption that ϕ lies in the Sobolev Space $H^1(\Omega)$, in our application these spaces are easily understood on an intuitive level and have a natural physical meaning. While doing the research, we were heading to two directions (patent code – [4]). In terms of research, we were defining the material magnetic properties while using magnetic space characteristics measurements. In the case of (patent code – [5]) research, we were defining electric characteristics on the surface with electric voltage/current values. Next, are going to discuss so-called electric impedance tomography EIT problems. There are going to be similar problems in the field of magnetic material characteristics. The inverse problem, as formulated by Calderon, is to recover σ from Λ_σ, for one set of Dirichlet and Neumann data, provided it contains enough frequency components, is enough to determine the boundary between two homogeneous materials with differing admittivity.

Electrical imaging system uses a system of conducting electrodes attached to the surface of the body under investigation. One can apply current or voltage to these electrodes and measure voltage or current respectively, measurement principle presented on Fig. 2.

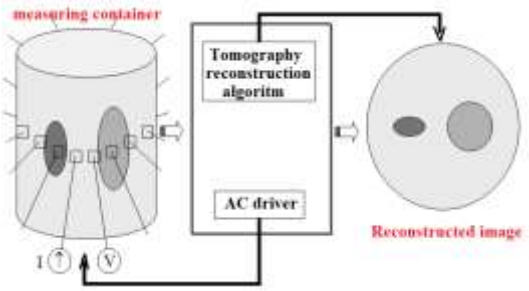


Fig. 2. The EIT tomography principles

The total variation functional is assuming an important role in the regularization of inverse problems in the image restoration context. The use of such a functional as a regularization penalty term allows the reconstruction of discontinuous profiles.

Newton-like algorithms is one of usual methods to solve inverse problem of (EIT), but it is sensitive to initial values. The PSO, genetic algorithm, Markov Chain Monte Carlo (MCMC) and many other methods for solving inverse problem are presented in literature. These methods are less sensitive for initial value but there are other problems that need to be resolved. The stable inversion method requires regularization. The efficient regularization is Tikhonov regularization. With applied the Tikhonov regulation a nonlinear total variation functional regularized inversion in a short time.

2.2. Hexapod Robot Basics

The precise modeling and controlling of driving DC motors are essential parts for the optimization procedure of the Szabad(ka)-II hexapod robot. The simulation model of Szabad(ka)-II hexapod was already built and validated [1]. It includes the detailed model of Faulhaber coreless DC micromotor and gear as engine of robot arms [1]. Model is implemented in Simulink environment, which forms the connection between the electrical and mechanical sides. The electrical side contains the model of embedded motor controller (PID or fuzzy controller), the model of sensors (encoder and current sensor, accelerometer, gyroscope) and the model of PWM amplifier. The mechanical side contains the all mechanical parts of the motor and gear system, its efficiency and the three dimensional kinematics and dynamics of the 18 DOF hexapod robot.

The PI and fuzzy-PI controllers have been already developed and optimized for this robot, where the objective function was the quality of robot walking. In this paper the objective function is independent from the robot walking; therefore it can be used for different purposes. The quality evaluation by the fitness or objective function focuses to reduce the high peaks and jerks in the motor current and torque and the robustness against the motor and load parameters.

We defined 17 rule in the core of fuzzy inference system, which are intended to decrease the speed error *dn(t)[rpm]* while hold down the current of motor *I(t)[A]* and especially the changing of current *dI(t)[A/s]*. We have tested the motor with constant speed rotations as desired input (*7000[rpm]*). In this test the gear is ignored yet. In the comparison three different control scenarios have been done in order to check the difference between them.

The quality or fitness evaluation of a controlled process is the key to be able to quantify the performance and optimize the system parameters. Genetic Algorithm (GA) was used for the optimization, because it has been already researched, and the convenient developed programs were available for this purpose. Fig. 3 shows the block diagram of this optimization procedure. For the equal opportunity competition the parameters were optimized both for PID and Fuzzy controller separately, while the input conditions, fitness evaluation and the optimization algorithm with its own parameters were the same. The details of the controller optimization, the selected fuzzy parameters to be optimized, and the setting of the optimization algorithm are not detailed here (in detail, see [1]).

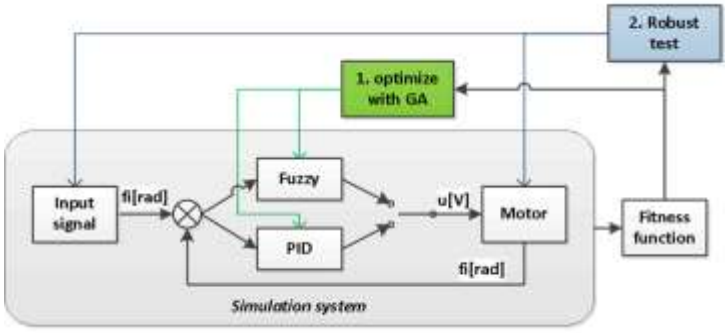


Fig. 3. Block diagram of optimization and robust test of motor controlling

The current dynamic model of hexapod walking - in view of character - is a multi-dimensional, highly nonlinear, non-smooth, and a slightly mixed integer problem, i.e., it has a minimum of seven dimensions.

PI controller has seven dimensions (five trajectories and two PI parameters), while the Fuzzy-PI has seventeen dimensions (five trajectories and twelve fuzzy parameters).

The system has non-continuous behavior due to walking on six legs and the ground contact; no random parts are included; contains integer parameters, e.g., the trajectory parameter order of filter is an integer type.

The ground contact model of six legs – a critical part of the dynamic model - has discontinuous characteristics. The applied fitness function is presented in [6]:

$$F = \frac{1}{\sqrt{\int q_{ERR}(t)^2\,dt}^{2} \cdot \sqrt{\int i(t)^2\,dt} \cdot \sqrt{\int \dot{q}_{ERR}(t)^2\,dt}^{0.5}}.$$

After the optimization of both controllers (see Fig. 4) the no-load and load tests demonstrated the difference between the PID and the fuzzy logic controllers.

Genetic algorithm (GA) – the figure shows the results of the bounded (limited) heuristic optimization method which was run on the parameters the Fuzzy engine regulator. Throughout the GA was executed with 31 generations and 80 populations (the top right graph), which all together 31x80=2480 resulted function calls. GA was searching for the maximum of the defined fitness function and the results are shown on the top left corner of the graph. On the right the fitness function values are indicated (red colored values are the higher ones).

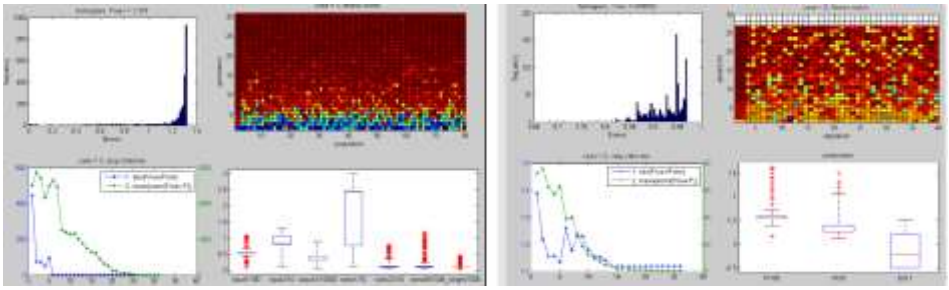


Fig 4. A) Fuzzy optimization; B) PID optimization

The bottom left figure shows the functions stopping criteria, which were built in the GA algorithm and was created by us. We can see that both values reached a near-zero value after the 31st generation. This is why the optimization stopped. The bottom right figure shows the optimization of the Fuzzy parameters' distribution for 2480 iteration. The values are normalized, they mean the following:

- Input1 – Fuzzy first input width
- Input 2 – Fuzzy second input width
- Output1 – Fuzzy output width
- Rulew1 – First fuzzy rule weight
- Rulew 23 - 2. and 3. fuzzy rule weight
- Rulew45 – 4. and 5. fuzzy rule weight

- Dfi_weight - angular velocity feedback

The figure shows a similar GA optimization for the PID values. Since there are less variables, the population is smaller (40), and the number of generations is 26. Three parameters have been optimized:

P – proportional member; I – integral member; D – derivative member

The consequence is that the fuzzy control strategy can keep motor the current at a significantly lower range compared to PID. The peak of motor current is higher than 5[A] in case of PID control, while fuzzy control keeps the current under 3[A].

3. Robustness Test

In our context, the robustness is the stability against the change of any system parameters including motor parameter or load parameters. Fig. 5 summarizes the analyzed parameters in this test. Each parameter was changed from 82% to 124% of its nominal value, while the change of fitness values is evaluated using relative standard deviation (RSD).

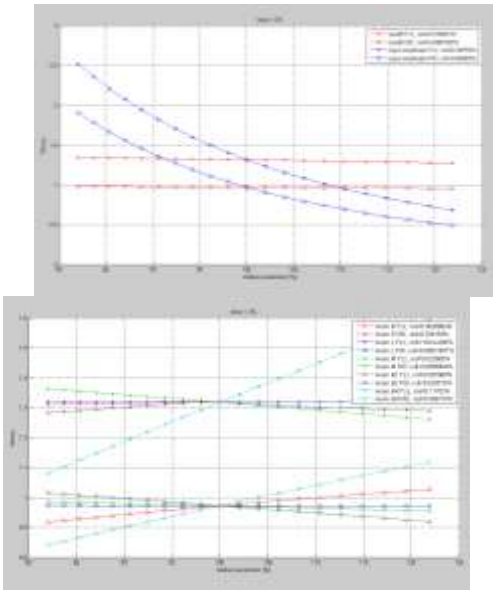


Fig. 5. System parameters analyzed in robustness test

We have chosen two stopping criteria, which can be seen in the bottom left figure:

- abs(Fmax/Fmin) - this shows the ratio between the largest and smallest function values (fitness value) in the current population. If the value approaches more and more (from the top) the value 1, then it converges towards the minimum.

- mean(norm(Fmax-F)) – this criterion is the average of the distance between the highest value in the current population (which is the best here) and the other function values. The value goes to zero if GA converges towards the optimum.

The boundaries of the variables were set based on experience. If it was realized that the values convergence towards extreme values then the boundaries were expanded. Usually, after 2-3 cycles, we accepted the result. In both cases, the resulting optimum for the test system has been verified and accepted as an optimum.

Two parameters were changed in the simulation system: load friction (load B, red) and the input speed's value (input amplitude, blue). The values have been changed from 84% to 122%, which is visible on the horizontal axis. The vertical axis shows the fitness value that changes as it is not normalized. However, from the robustness' point of view, the point is to change as little as possible, which means that the quality is less modified to any change in system parameter.

The degree of change is expressed by relative standard deviation (RSD), which is also shown in the legend in the figure. The PID and Fuzzy controllers are matched in the figure. The PID shows a more robust result of the load shift, and Fuzzy is the input's amplitude.

The figure shows the same as the previous one; only the engine parameters have been changed. It's more robust between the different engines. On average, Fuzzy resulted a bit more robust outcome, i.e., smaller deviation values (RSD). More details are published in [6]. Moreover, Fuzzy gave better robustness results because the deviation of fitness value (RSD) was smaller with 3% (7.91/8.14=0.97) on average compared to PID.

4. Conclusion

In this article, we represent two problems at inverse task model creation. In case of these systems, there are seemingly no common characteristics. But there is one common problem, which is the need to find one robust solution. There are two ways for finding the solution for robustness. The first option is when we know the whole description (in form of differential equation) of the examined system. In this case, the linear solution can give a good result. It is used a lot of times (in both problems) with bigger and smaller or success as well. We present not only the use of linear solution mechanism but the robot with examples, how does the heuristic solving mechanism work.

Another issue arises when we cannot describe the whole system, they are nonlinear (we do not know the nonlinear measure and type). In a lot of cases, there are systematic errors, which cannot be eliminated. The problem solvers are disturbed by high noise level, the noise itself and the correlation of system's parameters with noise.

Our experience shows that the linear solution works until a certain level, but by combining the heuristic solution, even more at the same time, it can help a lot in cases of making the result even better.

References

1. Kecskes, I., Burkus, E., Bazso, F., Odry, P.: Model validation of a hexapod walker robot. Robotica, Vol. 35, No. 2, 419-462. (2015).

2. Kecskes, I., Odry, P.: Optimization of PI and fuzzy-PI controllers on simulation model of Szabad(ka)-II walking robot. International Journal of Advanced Robotic Systems, Vol. 11, No. 11, 186. (2014).
3. Laslo Tadic, Miodrag Popovic, Peter Odry: Fuzzified Gabor Filter for License Plate Detection. Engineering Applications of Artificial Intelligence, Vol. 48, 40-58. (2016).
4. Magnetic tomography of ferromagnetic materials dept structure, PNo: 1600165
5. Data acquisition and processing method to assembly soft tomography examinations: PNo: 1500616
6. Kecskés, I., Burkus, E., Király, Z., Odry, Á., Odry, P.: Comparation of motor controllers using a simplified robot leg: PID vs fuzzy logic. In: Proceedings of the 4th International Conference on Mathematics and Computers in Sciences and Industry. Corfu, Greece. (2017).
7. Cariou, C., Berducat, M., Gobor, Z., Seiferth, B.: Mobile Robot Trajectory Planning Under Kinematic and Dynamic Constraints for Partial and Full Field Coverage. Journal of Field Robotics, Vol. 34, No. 7, 1297–1312. (2017).

Experiences in Teaching an IT Retraining Course – Fundamentals of Programming using Java

Dinu Dragan[1], Dušan B. Gajić[1], and Veljko B. Petrović[1]

[1] University of Novi Sad, Faculty of Technical Sciences
Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia
{dinud, dusan.gajic, pveljko}@uns.ac.rs

Abstract. The demand for IT experts in Novi Sand and, indeed, in Serbia has not only failed to decline, but has grown by leaps and bounds in the past few years. However, this demand is not followed in other job domains. As a result, a lot of unemployed people are interested in IT retraining, which is followed by increase in IT schools for retraining and retraining courses offered. In this paper the topic of teaching fundamentals of programming to adults with minimal or virtually no previous experience in computer programming is explored. We describe the motives, the structure, and outline of one such course based on the Java programming language. Java has been chosen because it is a baseline for many IT job positions and as the most direct way for producing junior programmers with at least minimal skills. We conclude the paper with a short report on the results that students achieved and an overview of the overall success of the course.

Keywords: IT Retraining, Education, Teaching Programming Basics, Introduction to Programming, Java.

1. Introduction

The course considering improvements in the field of information technologies (IT), set by the Serbian government in 2010 [1] has been steadily followed. Serbian economy and society continue to grow in the number of domestic IT companies and IT educated people. However, the struggle to retain as many as possible of the IT specialists in Serbia continues and only modest success in this sense has been achieved.

Number of IT companies in Serbia is constantly increasing [2]. Novi Sad, as the second largest city in the country, has a very prominent position in the Serbian IT industry (Fig. 1), but also in the world. As of 2015, there are more than 120 IT companies of various sizes in Novi Sad [3], which represents 6.7% of all Serbian IT companies. More than 5800 IT professionals work in Novi Sad, which is more than 50% of all Serbian IT professionals [3]. Three of the biggest Serbian IT companies are located in Novi Sad and together they employ more than 2300 people [3].

IT industry makes a huge potential for Serbian export growth as more than 90% of IT companies work for foreign markets [3]. Serbian IT market's value in 2015 was estimated to be around 450 million dollars [3]. One of the biggest barriers in further increase of the

Serbian IT industry size is a profound lack of IT professionals. The growing demand for IT workers has been identified by the government as a problem and it has been answered by making IT education available to more students. The number of students allowed to enroll in ICT studies within public faculties increased. The number of scholarships available to these students [4] increased also. At the Faculty of Technical Sciences, which is the primary educator of IT workers in Novi Sad, number of students allowed to enroll in ICT based studies is now around 1200 [5]. The studies involving software developing are the ones increased in student numbers the most [4].

Fig. 1. Trends in the Serbian IT industry – number of IT companies and IT workers situated in Novi Sad in regard to the rest of Serbia

The educational cycle of IT professionals is for many IT companies too long. From 3 to 5 five years are required for a faculty to produce a junior developer. Also, the best programmers usual end up either abroad or in the biggest companies that could offer the best working conditions. A number of small and medium companies are willing to employ less educated IT professionals [4]. This workforce is usually paid less, their training cycle is short, and expectations considering their knowledge are limited.

In contrast to the IT sector, there are a lot of unemployed people in the workforce that are willing to retrain and acquire certain IT skills. Experience shows that there are some non-IT profiles that can attain IT skills. These are usually similar engineer profiles or basic sciences such as physics and mathematics. As a result, several courses for IT retraining were created during the years. One such course, named Retraining for IT, was introduced by the Serbian Ministry of Public Administration and Local Self-government and the United Nations Development Program in Serbia [6].

Retraining for IT is a pilot project that aims to develop, in a very short timeframe, IT skills in people that had no previous experience or had minimal experience in IT. As the goal was to produce IT juniors is a short time period, the course was limited to 140 class hours that were spread over three months. So the challenge, besides the obvious ones, was to define a single programming language that would enable a more focused retraining course and enable the students to develop a full set of skills needed for a junior programmer. In the end, we chose Java as the underlying programming language due to several reasons. First and foremost, there is a huge demand for Java programmers in Novi Sad. Further, using Java enabled us to give a more focused retraining and we also considered its lack of pointers as an advantage. There are, of course, some disadvantages arising from fact that Java is an object-oriented language and teachers could get in temptation to start explaining definitions of classes, class methods, class instances, and

so forth, before they explain programming basics. All these issues are considered and elaborated upon in this paper.

The organization of the paper is as follows. In Section 2 we describe the terms such as training, retraining, and education. The entire Retraining for IT pilot project is also described in the same section. Different introduction to programming courses based on Java are compared in the final part of Section 2. Structure and content of Fundamentals of Programming course is described in Section 3. The results of the course and possible improvements of the course are described in Section 4. Section 5 concludes the paper.

2. Background

The term 'retraining' means teaching someone new skills that will enable them to do a different job [4]. Like the term 'training', it emphasizes *how* rather than *why* [7]. In contrast to education, it has a short learning cycle and it emphasizes practice. Rather than teaching all of the concepts, retraining is focused on a single or several simple concepts and students adopt these concepts using specific tools for a specific job.

Retraining for IT pilot project was conducted in three Serbian major cities: Belgrade, Novi Sad, and Nis. One hundred students took part in the project, 50 in Belgrade, 25 in Novi Sad, and 25 in Niš. They have been chosen from more than 8000 candidates, based on an introductory test, provided by the Ministry of Public Administration and Local Self-government. In each city, different organizations managed and conducted the courses. Authors of the paper conducted the courses in Novi Sad and the experiences described in the paper came from the Novi Sad part of the pilot project. The Novi Sad part of the project has been organized and managed by the Vojvodina ITC cluster [2].

Retraining for IT pilot project is not the first of its kind. A similar project was conducted also by the Serbian government and the Vojvodina ITC cluster, but financed by USAID in 2014 [4]. However, these two projects differ in time span and students' background. The first project lasted over 6 months and it included four modules: web developer, front end developer, mobile application developer, and database administrator. Each module contained two courses which were preceded with three fundamental courses: Fundamentals of Programming, Fundamentals of Object-oriented Programming, and Database Fundamentals. The Fundamentals of Programming course was meant to teach students fundamental programming skill and to train them in the C programming language. Experiences from that course have been used in preparing the course described in the paper.

48 students attended the courses. Although they had very different technical backgrounds, programming skills, age and gender, most of them had some experience and skills needed for the jobs they had been retrained. In the end, 75% students managed to acquire a job in the IT industry.

Retraining for IT project was defined with several restrictions in mind. Only the students with none or minimal technical knowledge and programming skills have been admitted. The overall time span for the courses has been limited to four months—three months for teaching and one month for practice work in one of the Novi Sad IT companies. The project's goal was to train Java/Web junior developers. The module was adopted based on the internal research made by the Vojvodina ICT cluster among the Novi Sad IT companies, in which majority of them defined that they needed junior Java

programmers with skills for Web-based programming. Also, current trends in software development clearly set Java as the most popular and most used programming language [8], emphasizing the growing need for Java developers. Also, from authors' personal experience from teaching academic courses, Web development skills are among the easiest ones to acquire.

Retraining for IT project consists of six courses:

- The Fundamentals of Programming,
- The Fundamentals of Object-oriented Programming,
- Database Fundamentals,
- Advanced Programming Techniques,
- Web Back-End Development,
- Software Development Principles and Methodologies.

At the Faculty of Technical Sciences, introduction to programming courses are usually based on the C programming language [9]. Although C has its advantages as a programming language for beginners—once you learn C, no other programming language is too hard to learn—due to very strict time constraint and the fact that all the rest of the courses are designed towards using Java, it was decided that the introductory course into programming—The Fundamentals of Programming—should also be in Java.

By avoiding C, several issues and general disadvantages of C as a beginning programming language were avoided. There was no need for working with pointers and no need for memory management. There was not enough time for detailed explanations of these subjects and it was very unlikely that the students could comprehend them.

Teaching introduction to programming course in Java is not a new approach [10, 11, 12]. When Java for this purpose, there are three approaches found in the literature [11]:

- The object-first approach which concentrates on object-oriented programming principles and focuses on objects and inheritance before introducing any of the procedural elements. The procedural elements are always introduced in the context of an object-oriented paradigm.
- The fundamental-first approach starts with the idea that all of the introductory programming concepts should be understood by students before moving to onto the specific technical features of the language such as the object-oriented aspects of Java.
- GUI first approach which illustrates the properties common to all Java classes by using Java Applets and Graphical User Interfaces (GUIs). Students are first used to develop GUI programs and through developing GUI they learn programming fundamentals and object-oriented programming.

The fundamental-first approach is the most commonly used. One of the main reasons for this is that the fundamental concepts—drawn from procedural programing—are simpler to grasp than the object-oriented programming concepts [13]. Another advantage of the fundamental-first approach is that students receive applicable fundamental knowledge which will enable them, if necessary, to shift to a new programming language and/or paradigm. As authors agree with the fundamental-first approach, it was adopted for the Fundamentals of Programming course.

3. The Course Description

The structure of the course is described on Fig. 2. Eclipse Integrated Development Environment (IDE) [14] has been used as a developing tool for the course. Microsoft PowerPoint slides were used for lecturing.

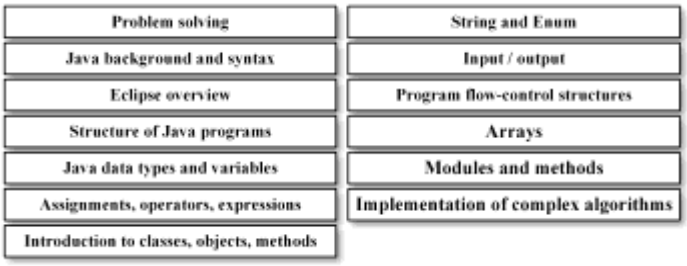


Fig. 2. The structure of the retraining course in programming fundamentals based on Java

As it can be observed in Fig. 2, almost all fundamentals concepts in programming have been included in the course. Each topic has been accompanied with adequate Java code examples. Teachers and students solved problems together using the Eclipse IDE. Besides the fully-developed examples, students were tasked with expanding or optimizing their solutions, or given an entirely new task for which solution could be adapted from the provided codebase. The final part of the course involved a test.

Creating the course in fundamentals of programming using Java was not without its challenges [4]. Students should be learning only programming basics unburdened with the object-oriented concepts of Java. The course that is following the Fundamentals of Programming, the Fundamentals of Object-oriented Programming, is also based on Java and it should teach the students the object-oriented concepts. However, Java is strictly an object-oriented language, meaning that it is not possible to clearly and absolutely separate fundamental concepts from object-oriented concepts and some basic introduction into object-oriented concepts has to be given. There are two approaches to performing this. The first approach is to teach the students that some things have to be done in a certain way, without further explanations. The second approach is to teach the students the minimal amount of concepts needed to understand object-oriented programming and then refer them to the next course for more details.

In our opinion, just teaching students to do things in a certain way—without explanation—is counterproductive, so the mix of the approaches has been adopted. For example, to create a new project in the Eclipse IDE, it is necessary to define a package and a class containing the main function which is static, Fig. 3. So, in the considered course, the hardest thing was to explain to students what static methods are without getting into too many details about object-oriented concepts such as classes, objects, and method ownership. Package and class can be explained as organizational units and most of the students understood them just like that. This was supported by the fact that, during the entire course, students did not have to create their own classes and they did not use more than one—main class—as a container for the rest of the code.

```
package prvi_projekat;
public class PrviProjekat {
    public static void main(String[] args) {
        System.out.println("Hello, World!");
    }
}
```

Fig. 3. The simple Hello World program in Java

Understanding objects has been a little trickier, but the need for objects arose only when String, Enum, and Arrays were introduced. As students earlier learned Java data types, it was easy to extend the concept of the class and to model the relationship between a class and an object relating to the relationship between a data type and a variable. The only difference was using the operator new which was explained as an operator for a simple memory management. Also, it was explained that objects, beside data, contain methods for manipulating data and that methods were called using the dot (.) operator. The explanation of the meaning of a static method has been left out in the beginning, because it was not essential at that moment. However, it is impossible to introduce fundamental concepts of code modularization in Java, Fig. 4, without introducing either the more detailed explanation of Java classes and objects or explaining the true meaning of static.

```
package first_project;
public class FirstProject {
    public static void main(String[] args) {
        say("Hello, World!");
    }
    public static void say(String message) {
        System.out.println(message);
    }
}
```

Fig. 4. Simulating function call in procedural programming language using the Java class method

As the goal of the subsequent course was to teach concepts of object-oriented programming, it was decided to explain to the students, in a simple way, what are static and non-static methods. For further explanation they were referred to the subsequent course.

4. The Course Outcomes

Unfortunately, it was observed that—as the course progressed—students' ability to follow the course and absorb the course material dropped significantly. This was most obvious when students were learning modularization and Java methods which confused them. Their confusion was cleared only when Java class methods were taught in the subsequent course.

The course contained 20 classes, taught over two weeks in blocks of 4 classes. As a result of very limited time, a large portion of the course was based on the expectation of

students exercising and practicing at home. At the beginning of the course, they were informed that in order for them to successfully finish the retraining courses and to absorb the thought material, they needed to practice at home and the time spent on practicing at home had to be at least two or three times longer than the overall length of the course. The students were highly motived, but, unfortunately, only 10% of the students actively worked at home. The reasons for the students' lack of practicing at home are various—students reported the pace of the course as being too fast for them or they had other private obligations.

At the end of the course, students were given a test. The test included 6 theoretical questions and a practical programming task. The practical task included the implementation of a simple program such as array sorting or search through array data. The test was graded with 30 points. The students scored an average of 19 points, which is 63% of the test maximum. It is interesting that they achieved far better results in theoretical part of the test, than in the practical part. Only 40% of students tried to solve the practical task. But the ones that tried mostly did it correctly. The overall success rate in this sense was 80%.

Although the indications before the tests were negative, the test results were actually better than expected. The reasons for this can be attributed to the following. Students had been given a week without classes to sort everything out and to absorb the material. Further, the difficulty of the test has been appropriately adjusted to their level of knowledge. The question remains whether they would achieve different results if they had more time for practicing and if the test was more practically oriented. For example, in previous iteration of the retraining course [4], students actually had been tasked with a single large project containing seven units including all the material taught at the course. They had two weeks to finish the project and, afterwards, they presented their work. At least four units had to be implemented successfully to pass the course. Teacher's expectations were that the results would be worse. But given the time span for course, the simpler test had to be accommodated.

5. Conclusion

In the paper the Java-based course in fundamentals of programming has been introduced. The course was designed to include as many fundamental programming concepts as possible by ignoring almost all of the object-oriented concepts which Java is naturally imposing. We described how we solved the cases in which we had to use some programming structures that Java programs had to have. Instead of completely ignoring packages, classes, objects, and static functions, we decided to give their simple explanations, which were later readdressed in a subsequent course which was teaching object-oriented concepts.

In this course, we learned that fundamentals of programming can be thought using the Java programming language, but with special care to some of the object-oriented concepts. In the end, students achieved better than expected test results. However, what the entire Retraining for IT project demonstrated is that the absolute beginners in programming need much more time for achieving basic skills in programming. Based on teachers' experience, at least as many classes are needed just for exercising and solving practical problems as for actually teaching the course material.

For future projects, we intend to increase the amount of classes and include more practicing. We would also like to involve students who had at least some contact with programming before. In the end we will compare the results from the modified course with the ones that students achieved in this course.

Acknowledgments

The reported research is partly supported by the Ministry of Education, Science, and Technological Development of the Republic of Serbia, projects TR32044 (2011-2017), ON174026 (2011-2017), III44006 (2011-2017), and III47003 (2011-2017).

References

1. Republic of Serbia, Službeni glasnik, Vol.51, 2010.
2. Web presentation of the Vojvodina ITC Cluster, available at: http://vojvodinaictcluster.org/, accessed on 14 July 2017.
3. M. Matijević and M. Šolaja: ICT in Serbia, At a Glance, Vojvodina ICT Cluster, 2015, available at: http://vojvodinaictcluster.org/wp-content/uploads/2014/08/ICT-in-Serbia-At-a-Glance-2015.pdf, accessed on 14 July 2017.
4. D. Dragan, V.B. Petrović, and D. Ivetić, Experiences in Teaching a Retraining Course in Fundamental Programming – Importance of Active Learning Strategies, in the Proceedings of the International Conference on Applied Internet and Information Technologies, 2014, pp.7-12.
5. Faculty of Technical Sciences, Enrollment documentation for the academic year 2017/2018, available at: http://www.ftn.uns.ac.rs/1966196897/konkurs-za-osnovne-akademske-studije, accessed on 14 July 2017.
6. Web-presentation of Retraining for IT project, 2017, available at: https://www.itobuke.rs/rs/, accessed on 14 July 2017.
7. M. Masadeh, Training, Education, Development and Learning: What is the Difference?, European Scientific Journal, Vol.8, No.10, p.62-68, 2012.
8. TIOBE Programing Languages Index for July 2017, available at: https://www.tiobe.com/tiobe-index/, accessed on 14 July 2017.
9. S. Mihić and D. Ivetić, A Simple Graphical Framework for the Acquisition 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. 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.
11. J. Yau and M.Joy, Introducing Java: the Case for Fundamentals-first, International Conference on Education and Information Systems: Technologies and Applications, 2004.
12. D.J. Eck, Introduction to Programming Using Java, 7th edition, 2014, available at: http://math.hws.edu/javanotes/.
13. I. Sanders and C. Mueller, A Fundamentals-based Curriculum for First Year Computer Science, SIGCSE Technical Symposium on Computer Science Education, 2000, pp. 227-231, 2000.
14. Eclipse IDE, 2017, available at: https://eclipse.org/ide/, accessed on 14 July 2017.

Semi-automated Interface Mapping in Enterprise Application Integration

Željko Vuković, Gordana Milosavljević

University of Novi Sad
Faculty of Technical Sciences
Chair of Informatics
{zeljkov, grist}@uns.ac.rs

Abstract. Enterprise information systems consist of software that covers a diverse set of needs and usually comes from many different vendors. In order to enable those individual applications to work together and free the data out of their respective information silos, enterprise information integration is performed. Syntactic approaches are mature, but require extensive knowledge about each of the involved systems. Semantic approaches use ontologies as a formal specification of the meaning of involved application interfaces. We present a hybrid solution that uses semantically annotated structural models to be semi-automatically mapped in order to generate code for an enterprise service bus.

Keywords: enterprise application integration, semantic integration, domain specific language, ontology, semantic conflicts, ESB, model-based

1. Introduction

Information needs of medium and large enterprises can seldom be sufficed by a single, monolith information system. Rather, these enterprises acquire individual applications, often from different vendors, each of which fits one scope of their needs. These can be: enterprise resource planning (ERP), product life cycle management (PLM), enterprise asset management (EAM), computer-aided design (CAD), materials requirements planning (MRP), computer-aided management and manufacturing (CAMM), etc. [1] These enterprise information systems are often characterized as HAD: heterogeneous, autonomous and distributed. Heterogeneity means that each application may implement its own data and process model and that the architecture (both hardware and software) it is built upon may be different from that of other applications in the enterprise. Applications are autonomous in the sense that each is developed to perform its task independent of others. The system is distributed because each application may run on a separate hardware system, usually reachable by network infrastructure. Some of the applications can themselves be described as distributed, since they consist of components that are located on separate networked computers.

Since each application is developed as autonomous, it would be unfeasible for developers of one application to foresee its possible interaction with all other involved systems. Consequently, the same data and processes become duplicated in many applications. As the applications themselves are unable to interact on their own, so called information silos are formed. To make changes of data in one application available to the others, we must integrate them. Other than synchronizing data changes across the system, integration can also enable the system to perform tasks that require functionality that is not available in any single application, but can be achieved by orchestrating their interaction.

There are many ways in which enterprise application integration (EAI) can be performed. One way in which they can be divided is to syntactic and semantic approaches. Examples of syntactic approaches are: (1) ad-hoc techniques – hard-coded solutions in a general purpose language, usually connecting a pair of applications at a time, resulting in o (n^2) complexity, (2) standard-based techniques – enforce that all applications adhere to a standard exchange format, (3) middleware-based techniques – data middleware such as ODBC, component-oriented middleware such as CORBA, or message-oriented middleware, such as Publish/subscribe messaging, (4) business process management techniques, and (5) service-oriented architectures (SOA).

The syntactic approaches are, by now, well defined, mature and with good platform and tool support. Their overview, mostly according to [1], is given in Fig. 1. In order to perform integration using any of these, extensive knowledge about each involved application is required. Since syntactic approaches look only at the structural model of data and processes, this knowledge must first be extracted, e.g. from documentation or by interview, and then applied to the integration solution. The system semantics are therefore contained in the solution itself, but they are not formally specified in any way. This means that (1) semantics cannot be formally verified and (2) cannot be reused.

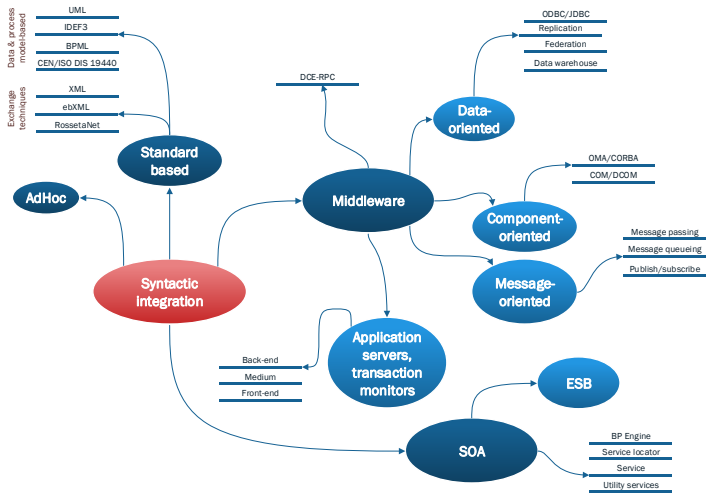


Fig. 1 Syntactic integration approaches

Semantic EAI approaches use ontologies [5] as a formal representation of knowledge [5]. Techniques and tools from the field of Semantic web enable the ontologies to be specified and persisted using languages such as RDF[1], OWL and OWL-S[2]. Tools known as reasoners can be used to deduce logical consequences from the set of facts contained in the ontology. The semantic description of the involved interfaces can be used to detect semantic conflicts among them. [6] One example of such conflicts are homonyms – interface elements from different applications that have the same name, but a different meaning.

Our solution is a hybrid of syntactic and semantic approaches. We define a framework that can be used to semantically enrich traditional structural models. The framework is then able to detect possible mapping candidates and semantic conflicts among them. Where automatic resolution of these conflicts is not possible, the user is informed of each conflict and can resolve it manually. Finally, code is generated that can be executed on a standard enterprise service bus (ESB).

2. The Framework

Our goal was to create a solution that can (a) reduce the time needed to integrate enterprise applications, (b) be testable in realistic industry scenarios and (c) be applicable to as wide range of scenarios as possible. We base our approach on existing languages and concepts for semantic description. Our approach lets the user annotate the structural model of the involved application with elements from an ontology. Together, they are used to automate the mapping and detect semantic conflicts.

A key point in our research was the decision not to develop a solution that should on its own cover *every* integration scenario in *any* context. Instead, we developed a framework that can easily be extended when a new set of challenges is identified in a particular scenario. Thus, the element mapping process has been divided into several phases, which the framework then executes. Each phase is implemented as a plug-in. By embedding our solution in a state-of-the-art ESB tool an executable prototype was attained, that enabled us to test the approach with an industrial partner in several real-word scenarios and benchmark its performance through experiments [2].

Our automatic mapping process is shown in Section 3. To find possible mappings, the framework iterates each output element of each output interface and examining the possibility of mapping it to each input element of each input interface. To determine if each of these pairs should be a mapping candidate, we test them against the available mapping criteria. These criteria can be defined by the user. Some examples of criteria that we have identified and used in our tests are given in Table 1.

[1] www.w3.org/TR/rdf-schema
[2] www.w3.org/Submission/OWL-S/

Table 1 Examples of mapping criteria. IE – interface element, OE – ontology element

Name	Description
Equal name	IEs have the same name (case ignored)
Equal XPath	Input and output IE are XML nodes and have the same XPath expression
Same annotation	IEs are annotated with the same OE
Aggregation	Two or more output interfaces are annotated with OEs that are `part_of`[3] an OE by which an output IE is annotated
Generalization	Input IE is annotated with an OE that is subclass of the OE annotated to the output IE.
Specialization	Output IE is annotated with an OE that is a subclass of the OE annotated to the input IE.
Splitter	Two or more output IEs are annotated with OE that are `part_of` an OE by which an input IE is annotated
Rejection	Input and output IEs are annotated by elements that are marked as `distinct` in the ontology

Once all the mapping candidates have been found, they are then inspected order to detect conflicts and attempt to resolve them. In this subsection we describe some of the conflicts that can be detected.

Multiple mappings for the same reason. If an output interface element was mapped with several input elements for the same reason, the user will be presented with a selection dialogue where they can choose which of the input elements should remain mapped. By *same reason* we mean that the mapping candidates are the result of the same criterion and that the reasons within that criterion are the same for all mappings. *Example*: let there be input interface elements I1 and I2, and an output interface element O1. They are all annotated with the ontology element A. Two mappings were made by the *Equal annotation* rule: I1 to O1 and I2 to O1. It is unclear if both I1 and I2 really have the same semantics and should be concatenated or have been annotated the same by mistake. Therefore, a warning is displayed and choice of selection is given.

Specialization ambiguity. When an output element is annotated with an ontology element that is more general than that of the input element, it cannot be known for sure if they should be mapped or not. The user is presented with a choice of making this mapping.

Type compatibility. Data types of interface elements are considered when building the output element expression. If the input type cannot be directly assigned to the output type, a conversion component is inserted automatically to make the mapping possible.

[3] As OWL does not have a built in way to represent part-whole relationships, we have used www.w3.org/2001/sw/BestPractices/OEP/SimplePartWhole

3. Implementation

We have implemented a proof of concept by modifying an open source integration platform - Talend Open Studio for ESB[4] (TOS). Open Studio is built upon Eclipse Rich Client Platform and consists of components communicating on an OSGi[5] plugin infrastructure. We have modified two of those components: tMap and tXMLMap. Both of these components have built-in Auto map feature, but that feature can only map elements that have the exact same name. However, this was a good starting point to implement an architecture for our framework. A simplified UML class diagram of the automatic mapper is given in Fig. 2. The input and output interfaces are tree structures. Each interface element (called *column* in TOS) is a `TreeNode`. Each `TreeNode` holds a list of ontology elements that annotate it, represented by instances of the Jena framework `Resource` class. Matching criteria described earlier are contained in classes that implement the Matcher interface. This is one of the extensibility point of the implementation, as other similar classes can be added as needed.

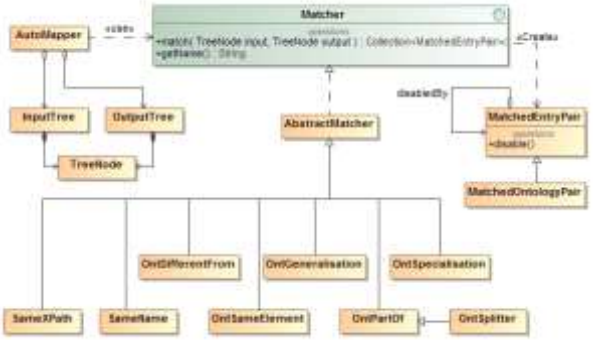


Fig. 2 UML Class diagram of the framework implementation

When the user clicks the Auto Map button in the tXMLMap or tMap dialog, the `map()` method is called from the `AutoMapper` class, which starts the process of running all the matching phases described earlier. It traverses all available input and output trees and for each input-output element pair it calls the `match()` method of all Matchers. The result is a collection of `MatchedEntryPair` instances. These pairs are then subject to conflict detection and resolution. When all conflicts have been resolved the user is presented with a graphical view[6] of mappings, as can be seen in Fig. 3.

[4] https://www.talend.com/products/application-integration/esb-open-studio
[5] https://www.osgi.org/
[6] The graphical representation of interface mappings is an existing TOS feature. In the same dialogue it is also possible to manually add or modify mappings using drag-and-drop.

Fig. 3 Talend Open Studio tXMLMap component showing mapping results

4. SAIL – a Domain-specific Language

In order to further simplify implementation of new mapping and conflict detection criteria, we have developed a domain-specific language (DSL), called SAIL [3]. This removes the need to use a general programming language like Java in order to extend the framework implementation. The DSL has been implemented using Eclipse Xtext. A simplified grammar of the language is given in Listing 1.

```
Model: elements+=Element*;
Element: Matcher | Conflict | ExpressionBuilder;

Matcher: 'Matcher' name=ID
(fullName=MatcherFullName)? condition=Condition
(continuation=ContinuationExp)?

Conflict: 'Conflict' name=ID matcher=ConflictMatcher
(condition=Condition)?
(causeMul=ConflictCauseMultiplicity)?
resolve=ConflictResolve;

ExpressionBuilder: 'OutExpression'
matcher=ConflictMatcher
(condition=Condition)? 'out' out=OutputExp;
```

Listing 1 A simplified grammar of SAIL language

Specification written in SAIL is automatically transformed into Java classes that implement architecture described in Section 3. An editor is provided for SAIL that supports code completion, syntax highlighting and validation, as shown in Fig 4.

Fig. 4 Example specification written in SAIL using the editor built for the DSL

5. Conclusion

We have presented a framework that can use semantically annotated structural model of application interfaces in order to find mapping between interface elements and detect semantic conflicts. The framework is extendable. Custom mapping and conflict detection criteria can be provided either directly, as an OSGi plugin developed in Java, or specified using SAIL DSL. The semantic description of an application's interface can later be reused in other integration scenarios.

Since manual interface mapping is a laborious task, this automation and reuse can speed up the development of and integration solution, as confirmed by our benchmark in [2]. Detection of semantic conflicts can prevent integration errors. This prevents loss of revenue for the enterprise. When integration is performed in a critical scenario where, e.g. medical equipment or patient data is involved, by preventing integration errors we can prevent loss of life.

Today, the need of integration is not confined only to enterprises. Any field that involves exchange of data or need for cooperation by systems that have not been explicitly designed to interoperate with each other (e.g. smart cities, cyber-physical systems, Industry 4.0) can benefit from techniques presented here.

6. References

1. Izza, Saïd. "Integration of industrial information systems: from syntactic to semantic integration approaches." *Enterprise Information Systems* 3.1 (2009): 1-57.
2. Vuković, Ž., Milanović, N., Vaderna, R., Dejanović, I., Milosavljević, G., Malbaša, V. "Semantic-aided automation of interface mapping in enterprise integration with conflict detection." Information Systems and e-Business Management 15.2 (2017): 305-322
3. Vuković, Ž., Milanović, N., Vaderna, R., Dejanović, I., Milosavljević, G. "SAIL: a domain-specific language for semantic-aided automation of interface mapping in enterprise integration." OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". Springer, Cham, 2015.
4. Guarino, Nicola. "Formal ontology and information systems." Proceedings of FOIS. Vol. 98. No. 1998. 1998.
5. Ontology-based integration of data sources. In: Information Fusion, 2007 10th international conference on, IEEE, pp 1–8
6. Goh CH (1996) Representing and reasoning about semantic conflicts in heterogeneous information systems. PhD thesis, Citeseer

On the Modeling and Simulation of M/M/1/k/k Queues with Stochastic Petri Nets

Ilija Hristoski[1]

[1] Faculty of Economics, Marksova 133,
7500 Prilep, Republic of Macedonia
ilija.hristoski@uklo.edu.mk

Abstract. Generally considered a branch of operations research, queueing theory has proven to deliver trustworthy results that are often used in the processes of business decision making, regarding the planning of resources needed to provide a certain type of a service. In this paper, the focus is put on a specific class of queueing systems that are commonly found in practice: M/M/1/k/k. Having minded the fact that basic queuing systems are usually analytically modeled and solved using Continuous Time Markov Chains (CTMCs), they can be also presented and evaluated using Stochastic Petri Nets (SPNs), in an equivalent and consistent way. Moreover, numerical simulation approach has been utilized to evaluate the resulting SPN model by exploiting TimeNET 4.3, a software environment dedicated to modeling and simulation of several classes of stochastic Petri Nets. The findings confirm that (1) queuing systems and networks can be successfully modeled and evaluated using stochastic Petri nets, whilst (2) the usage of dedicated simulation packages like TimeNET can reveal significant insights into the dynamics of the analyzed SPN model vis-à-vis its stationary and transient behavior in a time-efficient manner and at a high accuracy.

Keywords: M/M/1/k/k queue, Stochastic Petri Nets (SPNs), TimeNET, modeling and simulation.

1. Introduction

Queues are a constituent, yet an inevitable part of our everyday lives. They are omnipresent; they can be noticed at airports, banks, restaurants, supermarkets, offices, hospitals… wherever there are clients looking for a service provided by a resource. Nevertheless, the resources in the world we live in are rather limited and therefore, whenever there is a competition for them, queues occur. According to the queuing theory, the limited resources are dispensed by servers. Queues, i.e. waiting lines of requests posted by clients, are formed in front of the servers. Clients, i.e. the entities that require service, are said to 'arrive' at the service facility and place 'service demands' on the resource. On the other hand, resources are said to have 'a finite capacity', i.e. there is not an infinity of them, nor can they work (i.e. service the requests) infinitely fast. According to Kleinrock (1975), if either the arrival times or the size of the service demands are unpredictable, then "conflicts for the use of the resource will arise and queues of waiting customers will form." [1]

Knowing the basic system characteristics (number of servers, capacity of the waiting room/the buffer size, arrival rate, service rate, utilization of the servers, safety capacity), the ultimate goal of a queueing system analysis is to estimate the values of the most crucial performance measures, including: average waiting time in the queues, average time spent at the servers (average servicing time), total average time spent in the system, average queue length, average number of jobs/clients being served, as well as total average number of jobs/clients in the system. All of these help in the complex processes of capacity managing, i.e. in estimating how many resources/capacity is needed to keep the performance of a given queuing system within pre-defined boundaries, or to explore the sensitivity of recommendations to certain assumptions or changes.

The findings of the queueing theory are crucial because when it comes to real life, recent research has shown staggering facts. For instance, according to some estimation, Americans spend roughly 37 billion hours each year waiting in line [2]. A polling study carried out in Great Britain in 2009 has found consumers literally plagued and frustrated by queues since it turns out that an average adult wastes 5 hours and 35 minutes queueing each month, which is equivalent to 5 months, 2 weeks and 5 days of an average lifespan [3]. According to the same study, the length of time consumers have to spend waiting in a queue also severely affects their purchasing decisions, since whopping 88% of them gave up and left the shop without making a purchase. These and other similar research findings just point out the great importance of such a global and intense problem vis-à-vis its high impact on personal efficiency, business performances, and, ultimately, on the overall economy.

The paper is organized as follows. Section 2 deals with the most relevant and recent research regarding the computer modeling and simulation of queueing systems, in general. Section 3 elaborates the M/M/1/k/k queueing system. The class of Stochastic Petri Nets (SPNs) is briefly introduced in Section 4, whilst the modeling of M/M/1/k/k queueing system with SPN using TimeNET is presented in Section 5. Section 6 brings the performance evaluation, including both the stationary and transient analysis of the analyzed system, along with the corresponding simulation results. Section 7 concludes.

2. Related Research

During the last three decades, an abundance of research work has been done on simulating queueing networks.

For instance, Queueing Petri Nets (QPNs) combine the modeling power and expressiveness of both QNs and Generalized SPNs (GSPNs) into a unique class of Petri Nets that exhibit numerous advantages over ‘ordinary’ SPNs vis-à-vis the qualitative and quantitative analysis of systems [4]. In addition, QPN models allow the integration of hardware and software aspects of system behavior and lend themselves very well to modeling distributed systems. The excellence of QPNs in practice has led to the development of QPME (Queueing Petri net Modeling Environment), an open-source Java-based software tool for modeling and evaluation of QNs, based on the QPN formalism [5].

Hierarchical Queueing Petri Nets (HiQPNs), along with the software package HiQPN have been introduced in order to cope with hierarchical combinations of QPNs [6].

It should be also noteworthy to point out that numeric simulation of various classes of queueing systems has been also carried out by using general-purpose software and corresponding specific toolboxes. For instance, Petri Net Toolbox for MATLAB® has been used for modeling and evaluation of M/M/1 [7] and M/M/k (M/M/3) queueing systems [8], whilst SimEvents Toolbox for MATLAB® has been also used to convey performance analysis of a queueing network consisted of parallel M/M/1 queueing systems [9]. Nonetheless, dedicated software packages, like TimeNET [10] and SHARPE [11] have been also extensively used for modeling and performance analysis of various queuing systems and networks, e.g. M/M/1/k.

3. The M/M/1/k/k Queuing System

Taking into account the comprehensive notation for the classification of queueing systems, proposed by David George Kendall in 1953, in this paper the focus has been put on a specific queuing system, M/M/1/k/k, that can be commonly found in reality.

The M/M/1/k/k model is, in fact, a special case of the M/M/1/k model, known also as a 'finite capacity queue'. The latter one is an abbreviation of M/M/1/k/∞/FCFS, which assumes an existence of a single server, a limited buffer size, an infinite number of clients/jobs, and an FCFS (First-Come-First-Served) queuing discipline. The M/M/1/k/k model is therefore characterized by the following assumptions:

(1) Jobs/clients arrive according to a Poisson process with a parameter $\lambda \cdot t$, i.e. the time between two consecutive arrivals (i.e. inter-arrival times), *t*, has an exponential distribution with parameter λ. In other words, for $t \geq 0$, the probability density function (pdf) $f(t)$ is given by (1).

$$f(t) = \lambda \cdot e^{-\lambda \cdot t}. \quad \textbf{(1)}$$

(2) The service time, *s*, has an exponential distribution with a parameter μ, i.e. for $s \geq 0$, the probability density function (pdf) $g(s)$ is given by (2).

$$g(s) = \mu \cdot e^{-\mu \cdot s}. \quad \textbf{(2)}$$

(3) There is a single server, where the clients are processed in an FCFS (i.e. First-Come-First-Served) manner.
(4) The buffer, i.e. the waiting queue space has a limited size, *k*.
(5) The number of potential jobs/clients in the system is finite and equal to the waiting queue space, *k*.

At a first glance, it looks like there are no many real situations that could be possibly modeled as pure M/M/1/k/k queuing systems. This is because queues are always

associated with situations wherever there is concurrency among clients/jobs for resources, meaning that they outnumber the limited resources. However, there are many situations when the total number of clients/jobs in a system with a single server, at least temporarily, is less than, or equal to the number of available resources. That is, there is no explicit competition/conflict for resources since the available buffer size (i.e. the waiting queue space) is greater than, or equal to the population size (i.e. the clients/jobs). Real-world examples of such queueing systems can be found mostly in hospitality industry during out-of-season time periods, including food & beverage facilities, accommodation facilities, as well as travel & tourism facilities. Since all of these can be under-populated during a given time period/occasion, under the assumption that there is a single server, they can all be treated, equivalently, as M/M/1/k/k queuing systems by neglecting the possible surplus of available resources.

4. The Class of Stochastic Petri Nets (SPNs)

Originally proposed Petri Nets (PNs), also known as 'ordinary', 'non-timed', or simply, 'place/transition' Petri Nets, are recognized as a powerful tool for description and *qualitative* analysis of complex dynamic systems that exhibit features like concurrency, synchronization, and conflicts. The building blocks for constructing PN models include primitives like places, transitions, input and output arcs, and tokens.

The class of Stochastic Petri Nets (SPNs) represents an extension of PNs. They belong to the big family of *timed* Petri Nets, in which the basic model is augmented with time specifications, commonly used for *quantitative* analysis, i.e. performance and reliability evaluation of complex Discrete-Event Dynamic Systems (DEDSs). SPNs were introduced in 1980; they were deduced from the PNs by associating *exponentially* distributed firing times (delays) to all transitions in the model [12] [13] [14]. As such, they represent a convenient graphical and mathematical modeling formalism intended for describing/representing, as well as for performing both *qualitative* and *quantitative* analysis of DEDSs, whose dynamic behavior can be represented by means of homogeneous continuous-time Markov chains (CTMCs) [15]. In other words, there is a direct connection between SPNs and CTMCs in terms that reachability graphs of SPNs are equivalent to CTMCs, i.e. all SPN models can be converted to equivalent CTMCs, and vice-versa.

On the other hand, both the relationship and equivalence between Markov chains (i.e. CTMCs) and stochastic Petri Nets (i.e. Generalized SPNs and Stochastic Reward Nets) have been also extensively elaborated [11].

5. Modeling with TimeNET

The software package TimeNET (Timed Net Evaluation Tool)[1] is a general-purpose, graphical and interactive toolkit intended for modeling and evaluation of several classes

[1] The tool is free of charge for non-commercial use and is fully accessible for download from http://www.tu-ilmenau.de/sse/timenet/.

of stochastic Petri Nets including immediate, exponentially distributed, deterministic, or non-exponentially distributed (generally distributed) transition firing delays, as well as stochastic Colored Petri Net models [16] [17].

The M/M/1/k/k queuing system, modeled as an SPN, is depicted in Fig. 1. The structure of the model is quite simple. The place *P_idle* is a container of at most *k* clients/jobs, represented individually by *k* tokens, that are going to look for a service at some point in time. The number of clients/jobs is pre-defined, too, and stored within the variable *clients*. The value of the variable *arrival_time*, t_1, defines the mean inter-arrival time (measured in seconds, minutes, hours, etc.) of clients/jobs looking for service, arriving in the system at an arrival rate, λ. The value of the arrival rate λ is reciprocal of the inter-arrival time t_1, i.e. $\lambda = 1/t_1$. It is the parameter that defines the exponentially distributed firing delays of the transition *T_arrival*. Each firing of the transition *T_arrival* takes a single token from the place *P_idle* and puts it in the place *P_queue*, which resembles the real-life situation of a client/job entering the queue for service. The server itself consists of the place *P_queue* and the transition *T_service*. The server provides an exclusive service, i.e. only one client can use the service at the same time. The servicing of clients/jobs occurs, on average, every t_2 time units (seconds, minutes, hours, etc.). The value of the servicing time, t_2, is pre-determined, too, and stored within the variable *service_time*. It defines the servicing rate, μ, which is reciprocal of the time t_2 the server needs to serve a single client/job, i.e. $\mu = 1/t_2$. The parameter μ defines the exponentially distributed firing delays of the transition *T_service*. Each firing of the transition *T_service* takes a single token from the place *P_queue* and puts it in the place *P_idle*, which is equivalent to a real-life situation where a client/job leaves the servicing facility, after being serviced. All the arcs in the model have a multiplicity of 1, which guarantees that clients/jobs both arrive at the server and are serviced one-by-one (i.e. one at a time). It should be also pointed out that, in order to keep the system alive, $\lambda \leq \mu$, i.e. $t_1 \geq t_2$.

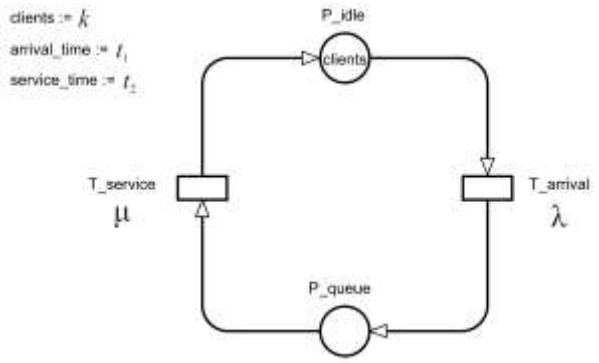


Fig. 1. A representation of an M/M/1/k/k queuing system as an SPN model in TimeNET [10]

This SPN model has been already described as an example of a Markovian M/M/1/k queuing system [10]. However, it becomes quite clear that the queuing system presented on Fig. 1 is not a pure M/M/1/k since it *a priori* assumes a *limited* capacity of

the system and *unlimited* population size. Rather than that, the SPN model depicts, in fact, a special case of M/M/1/k, namely, an M/M/1/k/k queuing system, since the (limited) capacity of the system, *k*, is identical to the (limited) size of the population, *k*. In other words, the proposed SPN model cannot and does not depict any real-life situation where the population size is infinite or even greater than the capacity of the system. In fact, only for large values of *k* (i.e. $k \to \infty$), the modeled system acts more like an M/M/1/k queuing system since the probability of having *k* clients/jobs waiting for a service in the place *P_queue* radically decreases with the increase of *k*. For $k = 1$, the SPN model in Fig. 1 can be utilized for investigating system availability [18].

6. Performance Evaluation with TimeNET

Performance evaluation of any SPN consists of two aspects: *stationary* (steady-state) analysis and *transient* (time-dependent) analysis. Both types of analyses have been carried out using the values of $t_1 = 60.0$ (*arrival_time*) and $t_2 = 50.0$ (*service_time*).

The structural analysis made with TimeNET shows that the number of reachable states is dependent on the number of clients in the system, *k*, and is equal to $k + 1$ (Fig. 2). There are no traps and no siphons within the model. The model contains a single P-invariant (*#P_idle + #P_queue = clients*); all places are covered by this P-invariant. It also contains a single T-invariant (*T_service*, *T_arrival*), which covers all transitions.

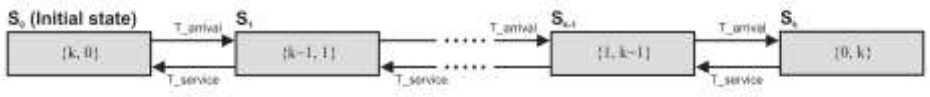


Fig. 2. Reachability/Coverability graph of an M/M/1/k/k queuing system. The marking of each state is given by the pair of values {*#P_idle*, *#P_queue*}, i.e. by the number of tokens within the places *P_idle* and *P_queue*, respectively. Shifts from one state to another are achieved by the firings of transitions *T_arrival* (at a rate of λ), and *T_service* (at a rate of μ). All states are tangible and since it is possible to go from every state to every state (not necessarily in one step), the resulting Markov chain is ergodic (i.e. irreducible)

The *steady-state* (i.e. stationary) evaluation assesses the mean system performance after all initial transient effects have passed, and a balanced operation mode (i.e. equilibrium) has been reached [16]. Fig. 3 offers a 3D view of the vectors of steady-state probabilities π [π_0 π_1 ... π_k], obtained for different number of clients/jobs in the system, *k* ($k = 1, 2, \ldots, 10$).

The *transient* (i.e. time-dependent) analysis of various types of stochastic PNs with TimeNET is based on the generation of a reduced reachability graph and the usage of the method of supplementary variables, showing graphically the evolution of specific performance measures from the initial marking (i.e. time zero) up to a given end time (i.e. transient time), by analyzing the model behavior [16]. This kind of analysis is always performed in order to evaluate the dynamics of a set of pre-defined performance measures over time, including Utilization, Average queue length, Average waiting time, and Probability of having *x* ($x = 1, 2, 3, \ldots, k$) clients/jobs in the system, both waiting in

a line and being serviced (Table 1). The results of the transient analysis vis-à-vis these performance measures are depicted in Fig. 4, Fig. 5, Fig. 6, and Fig. 7, respectively.

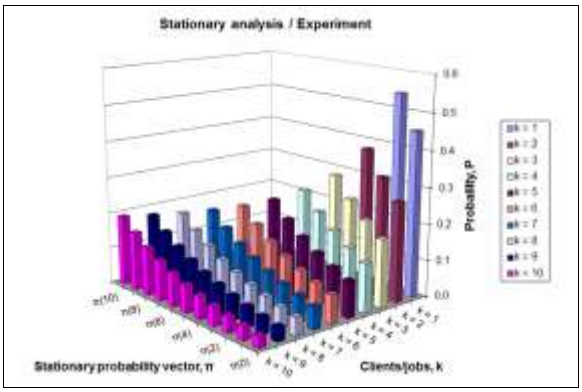


Fig. 3. Graphical representation of the family of vectors containing steady-state probabilities, π [π_0 π_1 ... π_k], obtained for different number of clients/jobs in the system, k (k = 1, 2, …, 10) during the stationary analysis of the SPN model (λ = 1/60; μ = 1/50) (Author's own calculations)

Table 1. The definition of crucial performance measures using the TimeNET syntax

Performance measure	TimeNET definition
Utilization	P{#*P_queue* > 0}
Average_queue_length	E{#*P_queue*}
Average_waiting_time	E{#*P_queue*}/(1/*arrival_time*)
Prob_*x*_clients_in_system	P{#*P_queue* = *x*}

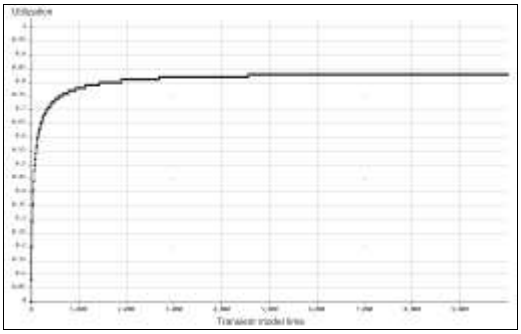


Fig. 4. Graphical representation of the Utilization, obtained by the transient analysis of the SPN model (k = 1,000; λ = 1/60; μ = 1/50; Transient model time T = 10,000), as generated by TimeNET (Author's own calculations)

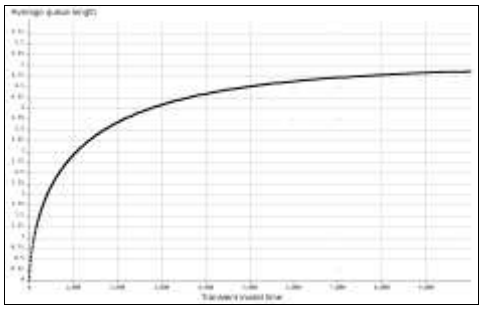


Fig. 5. Graphical representation of the Average queue length, obtained by the transient analysis of the SPN model (k = 1,000; λ = 1/60; μ = 1/50; Transient model time T = 10,000), as generated by TimeNET (Author's own calculations)

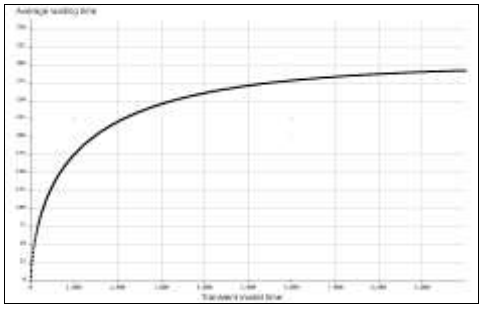


Fig. 6. Graphical representation of the Average waiting time, obtained by the transient analysis of the SPN model (k = 1,000; λ = 1/60; μ = 1/50; Transient model time T = 10,000), as generated by TimeNET (Author's own calculations)

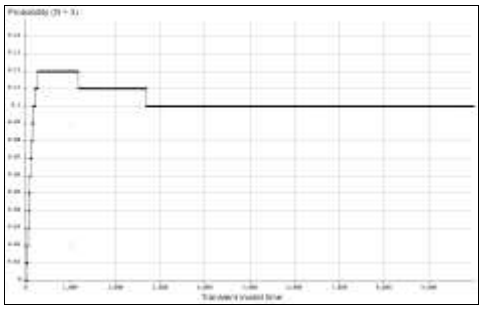


Fig. 7. Graphical representation of the Probability of having N = 3 clients/jobs in the system, obtained by the transient analysis of the SPN model (k = 1,000; λ = 1/60; μ = 1/50; Transient model time T = 10,000), as generated by TimeNET (Author's own calculations)

In addition, Fig. 8 shows the non-linear dependencies of all performance measures on the inter-arrival time, defined by the variable *arrival_time*, t_1.

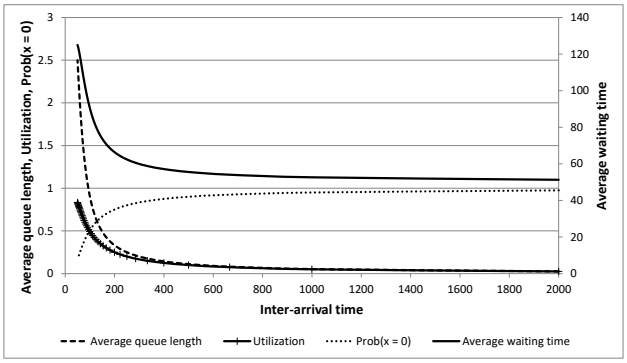


Fig. 8. Graphical representation of various performance measures' dependencies on the inter-arrival time $t_1 \in [50, \ldots, 2{,}000]$ ($k = 5$; $\lambda = 1/t_1$; $\mu = 1/50$) (Author's own calculations)

7. Conclusion

The paper provides an evidence for the stochastic modeling and performance evaluation of one of the simplest Markovian queuing systems, the M/M/1/k/k, by employing the class of SPNs. The usage of formal models like those implemented as SPNs has a number of advantages over writing simulation programs from scratch: not only it is much easier to set up a simulation that is free of logical errors, but also various qualitative properties can be verified against the modeled system far more easily for a formal model than for an *ad hoc* computer simulation program. In that context, the support offered by TimeNET, dedicated software providing a superb modeling and simulation environment oriented towards the evaluation of various classes of stochastic Petri nets, is both highly welcomed and recommended. The findings of the performance evaluation conducted here confirm that (1) queuing systems and networks can be successfully modeled and evaluated using stochastic Petri nets, whilst (2) the usage of dedicated simulation packages like TimeNET can reveal significant insights into the dynamics of the analyzed SPN model vis-à-vis its stationary and transient behavior in a time-efficient manner and at a high level of accuracy.

References

1. Kleinrock, L.: Queueing Systems, Volume I: Theory. John Wiley & Sons, New York, NY, USA. (1975)
2. Stone, A.: Why Waiting Is Torture. Sunday Review, The New York Times, New York (2012). [Online]. Available: http://www.nytimes.com/2012/08/19/opinion/sunday/why-waiting-in-line-is-torture.html (current May 2017)

3. The Telegraph: Britons spend six months queueing. The Telegraph, London (2009). [Online]. Available: http://www.telegraph.co.uk/news/newstopics/howaboutthat/5052956/Britons-spend-six-months-queueing.html (current May 2017)
4. Bause, F.: Queueing Petri Nets: A Formalism for the Combined Qualitative and Quantitative Analysis of Systems. In Proceedings of the 5th International Workshop on Petri Nets and Performance Models. Toulouse, France, 14–23. (1993)
5. QPME: QPME Project Website. Die Julius-Maximilians-Universität (JMU), Würzburg, Deutschland (–). [Online]. Available: http://se.informatik.uni-wuerzburg.de/tools/qpme/ (current May 2017)
6. Bause, F., Buchholz, P., Kemper, P.: QPN-Tool for the Specification and Analysis of Hierarchically Combined Petri Nets. In: Beilner, H., Bause, F. (eds.): Quantitative Evaluation of Computing and Communication Systems. Lecture Notes in Computer Science, Vol. 977. Springer-Verlag Berlin Heidelberg, 224–238. (1995)
7. Kamceva, E., Tancev, F.: Modeling and simulation of queue waiting through the concept of Petri Nets. Lex et Scientia International Journal, No. XVIII, Vol. 1, 303–309. (2011)
8. Gacovski, Z., Kamceva, E.: M/M/k Queues Modelled by Using of Petri Net Simulator. In Proceedings of the XLVI International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST 2011), Volume 1. Niš, Serbia, 123–126. (2011)
9. Nuredini, R., Gacovski, Z., Ramadani, J.: Analytical and Simulation Performance Result Analysis for Parallel M/M/1 Queueing System. In Proceedings of the XLVI International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST 2011), Volume 3. Niš, Serbia, 783–784. (2011)
10. Hellfritsch, C.: TimeNet - Examples of Extended Deterministic and Stochastic Petri Nets. Technische Universität Ilmenau, Ilmenau, Deutschland. (2009). [Online]. Available at https://www.tu-ilmenau.de/fileadmin/media/sse/Software/TimeNET/Example_Models/eDSPN-examples.pdf (Current May 2017)
11. Bolch, G., Greiner, S., de Meer, H., Trivedi, K. S.: Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications, 2nd Edition. John Wiley & Sons, Hoboken, New Jersey, NJ, USA. (2006)
12. Symons, F. J. W.: Introduction to Numerical Petri Nets, a General Graphical Model of Concurrent Processing Systems. Australian Telecommunications Research, Vol. 14, No. 1, 28–32 (1980)
13. Natkin, S.: Les réseaux de Petri stochastiques et leur application a l'evaluation des systemes informatiques. Thèse de Docteur Ingegneur, CNAM, Paris, France. (1980)
14. Molloy, M. K.: On the Integration of Delay and Throughput Measures in Distributed Processing Models. Ph.D. Thesis, UCLA, Los Angeles, CA, USA. (1981)
15. Balbo, G.: Introduction to Stochastic Petri Nets. In: Brinksma, E., Hermanns, H., Katoen, J-P. (eds.): Lectures on Formal Methods and Performance Analysis. Lecture Notes in Computer Science, Vol. 2090. Springer-Verlag Berlin Heidelberg, 84–155. (2001)
16. Zimmermann, A., Knoke, M.: TimeNET 4.0: A Software Tool for the Performability Evaluation with Stochastic and Colored Petri Nets, User Manual. Technische Universität Berlin, Berlin, Deutschland. (2007) [Online]. Available at http://www2.tu-ilmenau.de/sse_file/TimeNET/ Documentation/TimeNET-UserManual40.pdf (Current May 2017)
17. Zimmermann, A.: Modeling and Evaluation of Stochastic Petri Nets With TimeNET 4.1. In Proceedings of the 6th International ICST Conference on Performance Evaluation Methodologies and Tools, Cargèse, France, 54–63. (2012)
18. Hristoski, I. S., Mitrevski, P. J., Kotevski, Z. G.: Stochastic Modeling of e-Commerce System's Availability. In Proceedings of the 3rd International Conference on Applied Internet and Information Technologies (AIIT 2014), Zrenjanin, Serbia, 21–26. (2014)

Entrepreneurial Environment and New Technologies as Development Factors of Competitiveness of the Domestic Economy

Dragan Ćoćkalo[1], Dejan Đorđević[2], Milan Nikolić[2], Edit Terek[2], and Sanja Stanisavljev[2]

[1]University of Novi Sad, Technical faculty "Mihajlo Pupin", Đure Đakovića bb, 23000 Zrenjanin, Republic of Serbia
dragan.cockalo@tfzr.rs
[2]University of Novi Sad, Technical faculty "Mihajlo Pupin", Đure Đakovića bb, 23000 Zrenjanin, Republic of Serbia

Abstract. Unemployment, especially among the young population, represents one of the crucial economic problems. In overcoming this problem of unemployment, all relevant institutions must get involved, in order to stimulate young people towards participating in entrepreneurial processes. Below it will be illustrated an analysis of attitudes of young people in the Banat Region regarding entrepreneurship and starting their own business. Within this research, carried out at the end of 2016 on the sample of 350 examinees, certain data was received about the opinions of students from the universities and higher education institutions oriented towards management and business and the key indicators were analyzed as well. Generally, the research indicates that young people are not stimulated for participating in entrepreneurial processes. Furthermore, the research shows that creating an appropriate environment, as well as the procurement of new technologies, are among the key factors for the development of the competitiveness of the domestic economy.

Keywords: entrepreneurial environment, new technologies, young as entrepreneurs, regional development.

1. Introduction

Creating new job positions, according to the experts from the World Bank (WB), is a necessary step in the development of national economy. Namely, in the report given by the experts from the WB it is stated that it is necessary to create 600 million new job positions until 2020, in order to maintain the economic development, employ young people and decrease potential unrest. Moreover, the demographic structure of the society represents a serious issue, because it has an influence on labor productivity. In some world countries there is a decline in the birth rate which affects the level of working age population for a longer period.

In the World Bank report it is claimed that, in order to return people to work by increasing the number of jobs, countries need to re-establish a momentum for economic

and institutional reforms which existed before the crisis so they could: a) establish the foundations for creating jobs for the workers by pushing the reforms in direction of making a friendly environment for the growth of existing companies, increasing their productivity or entering the market in order to activate entrepreneurial potential for the appearance of new companies that will succeed or fail fast without making great costs, b) carry out the policies that will support the workers so they can be prepared to do newly made jobs by showing their skills, initiative and unimpeded access to work and to be ready to be transferred to the positions with the greatest potential for creating jobs.

Rapid technological changes, require a new philosophical approach in studying corporate governance in terms of continuous monitoring and long-term understanding of technology development, which represents a constant dynamic process, as well as its prediction in the future in order to harmonize market potentials and marketing efforts, reduce risk, and exploit all positive sides of companies and chances in the environment. Enterprise management becomes information management.

In the 1960s, Peter Draker [1] formulated the phrase "era of discontinuity", by which he wanted to reflect the multiple development of science, technology and knowledge, international economics and socio-political ideologies. This discontinuity continued throughout the 1980s, except that the world became "smaller" due to developments in communication technologies. Today`s new stage of development Draker marks as a knowledge society.

There are three key determinants of technological progress [2]:

- the improvement of the existing and the creation of new technology,
- improvement of existing and creation of new products/services,
- improvement of organization and management, with application of information and communication technologies.

There are two emerging forms of technological advancement [2] - tangible and intangible technological advancement. Tangible involves the advancement of hardware components of technology, the introduction of specialized news into technological systems in the form of new equipment, tools, machines, new materials, physical objects, which promote production and business. Intangible technological advancement means improving the intangible components of technology: knowledge and skills of employees, software, organization and management.

The process of production and application of high technology is a highly scientific and professional work that requires the application of multi-criteria optimization due to the need of considering and analyzing a large number of parameters, which are often contradictory, from different aspects: technical, technological, economic, ecological, political and cultural.

New technology will in any case be accepted more quickly if resistance to tradition is less. Significant changes are hard for companies as well as for people. However, technological change can not be overcome, so many industries or companies are forced to adapt to changes in technology, and if they do not, they may be at risk of experiencing weakened market positions and sometimes disappearance.

In an information society, information becomes a strategic source of wealth, along with natural resources. Who possesses information (knowledge), possesses power. Communication is the core of the information society. The development of information

technologies has completely changed the communication process. As a basic consequence, there is a faster flow of information through the communication channel, which leads to closeness of sender and the receiver, which then leads to the abolition of the "information floating".

The most important technological breakthrough in marketing, especially in international marketing, was achieved in the field of market communication. The rapid development of information technologies and direct marketing had the most significant impact on the formation of a new marketing communication model, which is designated as integrated marketing communication.

It is clear that companies must fully follow all the trends in the field of high technology and communication infrastructure, and use them in order to make the most efficient way to reach consumers.

2. Analysis of the competitiveness of the domestic economy

Domestic companies have been uncompetitive internationally for many years. Low productivity and insufficient investment in achieving business quality are the main reasons for poor competitiveness of domestic enterprises. The inadequate competitiveness of domestic enterprises can be seen the most within the framework of the global economic crisis.

According to the World Economic Forum for 2016, Serbia was ranked as 90th out of 138 analyzed countries (Table 1). This represents a step forward in relation to 2015, when Serbia was ranked at the 94th place. Interestingly, this year Serbia has found itself accompanied with Honduras (88th place), Cambodia (89th place), Ecuador (91st place) and Dominican Republic (92nd place).

Table 1. Ranking of the Western Balkan countries towards competitiveness in the period 2012-2016. [3]

Country	Place in 2012.	Place in 2013.	Place in 2014.	Place in 2015.	Place in 2016.
Slovenia	56	62	70	59	56
Montenegro	72	67	67	70	82
Croatia	81	75	77	77	74
Macedonia, FYR	80	74	63	60	68
Serbia	95	101	94	94	90
B and H	88	87	-	111	107

From the countries near by, Austria is in the 19th place, Hungary is 69th, Bulgaria is 50th, Romania is 62nd, Greece is 86th and Albania is at 80th place. The basic requirements of the country's competitiveness consist of the following indicators: institutions, infrastructure, macroeconomic environment, health and basic education, higher education and training, efficiency of the commodity market, labor market efficiency, capital market efficiency, technological capability and market size (Table 2). In terms of the economy of the Republic of Serbia, the most problematic factors for

business are according to the Report of the World Economic Forum: the tax rates - 11.7%, inefficient state apparatus - 10.6%, political instability - 9.0%, access to sources of financing - 9.0%, corruption - 8.7 %, tax regulations - 8.1%, inadequately trained workforce - 7.9%, inadequate infrastructure - 6.9%, government instability - 6.3%, insufficient working ethics - 5.5%.

Table 2. Ranking of the Western Balkan countries according to the basic requirements of competitiveness in 2016 [3]

Country	Place in 2014. summary	Place according to basic requirements	Ranking according to business efficiency	Ranking according to innovation
Slovenia	56	38	54	37
Montenegro	82	86	79	98
Croatia	74	68	68	92
Macedonia, FYR	68	66	73	64
Serbia	90	87	90	120
B and H	107	94	106	122

According to the World Economic Forum, all countries in the world are divided into five groups. The first group consists of countries that base their economy on resources, the third group consists of countries that base their economy on the development of business efficiency, and the fifth group consists of countries that base their economy on the application and development of innovative activities (mainly the most developed countries in the world). The second group of countries consists of countries in transition from the first to the third group the fourth group consists of countries in transition from the third to fifth group of countries. Serbia is in the third group of countries (development of business efficiency), along with Montenegro, Romania, Bulgaria, Macedonia, B and H, Albania, Indonesia, China, South Africa, etc. Croatia is, for example, located in the fourth group, which includes countries such as Argentina, Brazil, Chile, Hungary, Latvia, Mexico, Poland, Turkey, etc., while Slovenia is in the fifth group of countries.

According to some indicators [4], productivity in Serbia is still 42% of the European average. The reason for the poor productivity of business in domestic enterprises lies in bad organization of work, technological backwardness and lack of knowledge. There are several reasons for the low productivity of the domestic economy. State-owned companies, which for political reasons did not conduct restructuring, have too many employees. Certain privatized companies also did not undergo the process of transformation because the subjects of privatization have been shaken by stronger turnarounds. Finally, there are also low investments. The Serbian industry is based on labor-intensive branches.

Outdated technology, poor quality, unattractive packaging and high prices are the main reasons why Serbian products can not be supreme. The smallest competition is at the processing industry, metal production and electronics, where for years there is no technological renewal. Business people believe, there should be customs and tax relief, reduction of state taxes, as well as reduction of prices of electricity, gas and fuel in order to increase competitiveness. It is necessary to raise the level of technological

equipment, because the average age of machines in Serbia is 30 years. In relation to the region, there is a delay of 12 years [5].

3. Methodology

The survey "The analysis of attitudes and opinions of young population in relation to starting their own business and socially responsible business performance" had for its main objectives to determine the attitudes and opinions of examinees on starting their own business, in other words, on the success of business practice in Serbian enterprises. The participants/examinees of the survey were the students of universities and high schools in Serbia (average age of examinees was about 22), oriented to business and management. The survey was carried out at the end of year, seven years in succession (2010-2016) in the following cities, towns and municipalities: Belgrade, Pančevo, Aleksadrovac, Bačka Palanka, Požarevac, Vršac, Paraćin, Jagodina, Smederevo, Novi Sad, Alibunar, Ivanjica, Nova Varoš, Čačak, Zrenjenin, Kraljevo. The survey was carried out by anonymous, structured questionnaire. The results were published on several occasions [6, 7, 8, 9]. The total number of examinees is nearly 4000, and, for example, in 2015 survey, 616 examinees were included.

According to the territory on which the results were collected, since 2013 we have also monitored and analyzed the results regarding the place where the examinees studied. These places were grouped in three regions – Vojvodina, Belgrade and Serbia without the provinces (in the results it was named South Serbia). The data from the collected questionnaire were processed and grouped regionally. This paper gives the comparative review of results obtained from the basic and control groups of examinees from the Central Banat region (Zrenjanin). It is a base of this research within the framework of the projects "Improving entrepreneurial climate, the analysis of aspects and possible courses of action among youth in the Central Banat region" funded by Provincial Secretariat for Higher Education and Scientific Research, Autonomous Province of Vojvodina. The sample included 350 examinees – 300 active students (200 from Technical Faculty "Mihajlo Pupin", 100 students from the Technical College of Professional Studies) and 50 examinees of the appropriate age who already finished their studies and who are currently unemployed, as a control group. The first results were reported in the paper "Development of the entrepreneurial behavior of young people - Research results from the Banat region" [10].

4. Analysis of Research Results Among Students

4.1. Entrepreneurial ambience

Generally and locally, the research showed that the majority of examinees looked affirmatively to private business as well as to starting and doing business. The biggest

number of examinees wanted to start their own business: basic group – 79.7%, control group - 80%. Private business is for the examinees equally "risk and uncertainty" (basic – 22.3%, control – 17.7%) and "challenge" (basic – 30.8%, control - 31%). They agreed (basic – 45.5%, control - 22%) or mostly agreed (basic – 35.6%, control - 56%) when asked the question "Do you think that people in our country still do not know legal and business possibilities of private enterprises?" They mostly agreed (basic – 48.7%, control – 55.1%) when asked the question "Do you think that private business is more successful than business in other forms of ownership?"

When speaking about the ambience for business start-up, the young (basic- 83.53%, control - 74%) mostly consider that the ambience is not encouraging and that the state has to be the key initiator of this process (basic – 89.3%, control - 96%).

Out of the suggested answers on the question “What are, in your opinion, the biggest limitations for your own business start-up?”, seen from the level of the total sample, three answers were distinguished: lack of financial means, unstable political and economic situation and too high taxes – Figure 1.

The data in Figure 2 show the necessary encouraging measures for young population for their business start-up. Three key supporting measures are distinguished: more favourable loans, education and laws/regulations related to young entrepreneurs. Market regulation is also high ranked among these measures.

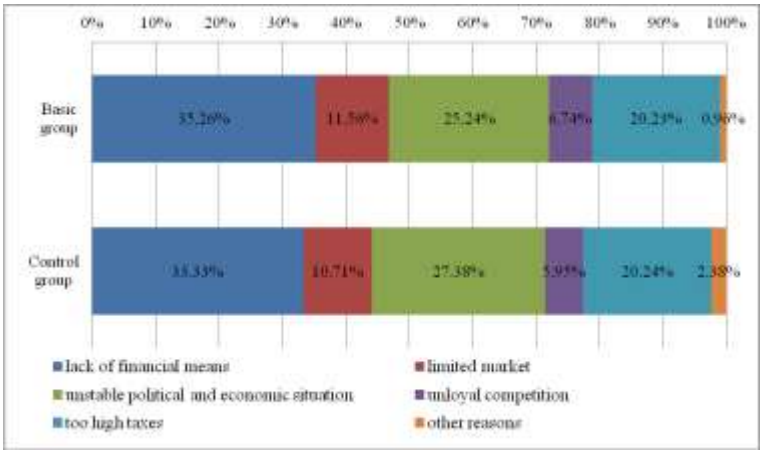


Fig. 1. Biggest limitations for business start-up

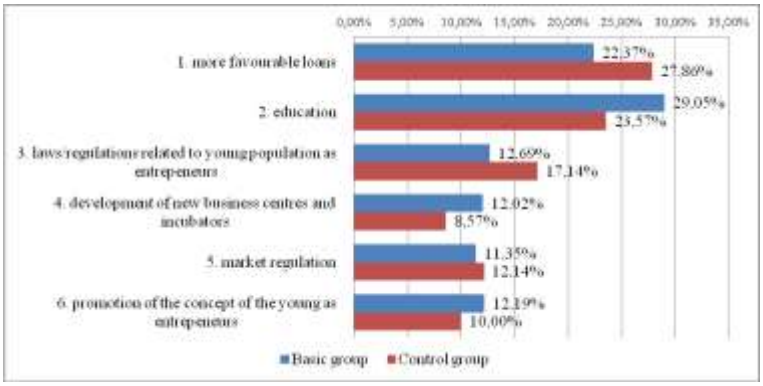


Fig. 2. Stimulating measures for business start-up

The aforementioned should serve as a starting point for the state institutions as well as the institutions on the local level/municipalities which deal with the issues of young population and economy in order to find solutions and create the ambience favorable for encouraging the young as entrepeneurs and their own business start-up.

4.2. Competitiveness of domestic enterprises

When assessing the competitiveness of domestic enterprises, 28.1% respondents of the basic group and 34.09% respondents of the control group believe that the competitiveness of domestic companies does not meet the requirements set by the international environment, 52.19% of the basic and 54.55% of the control group believe that the competitiveness of domestic enterprises partially meets the requirements that sets the international environment, while 14.23% of the respondents of the basic and 9.09% of the control group believe that domestic enterprises meet the conditions for achieving competitiveness in international frames, ie 5.47% respondents of the basic group and 2.27% of the control group consider it to be very satisfactory.

The most important factors that are missing in the development of the competitiveness of domestic enterprises are (Table 3): new technologies, education, employee motivation and financial support. When it comes to assessing the degree of innovation of domestic enterprises, the majority of respondents consider that domestic companies partially fulfill this factor of competitiveness (basic - 62.73%, control - 73.91%), ie they do not fully meet this factor of competitiveness (basic - 21.03% 19.57%), while only 2.58% respondents of the basic group consider that domestic companies are very innovative.

Respondents, as the necessary elements for the development of the competitiveness of domestic enterprises cited (Figure 3): application of modern methods and techniques of management, investment in the development of national brands, investments in marketing and purchase of new technology.

Table 3. Ranking of the Western Balkan countries according to the basic requirements of competitiveness in 2016 [3]

	Basic group	Control group
Implementation of QMS	2.40%	4.32%
Education	13.43%	13.51%
Marketing	7.62%	5.41%
Financial support	10.22%	10.27%
Innovation	8.22%	6.49%
New technologies	12.42%	13.51%
Motivation of employees	10.42%	11.89%
Cluster development	1.60%	1.08%
Prices	6.61%	8.11%
Payment ability of the buyer	4.81%	4.32%
Institutional support from the state	6.81%	5.95%
Appropriate business environment	5.01%	1.62%
Flexibility	5.41%	5.41%
Lack of professional workforce	4.81%	7.03%

Fig. 3. Necessary elements for developing competitiveness of domestic enterprises

5. Conclusions

The main problem of domestic enterprises, however, is the lack of competitiveness. Without the application of new knowledge, and the introduction of new technologies, domestic companies can not count on success at foreign markets. The business of domestic enterprises must be based on the application of management techniques that support competitiveness, innovation and flexibility, as well as on the interventional improvement of knowledge of employees in them, especially of the executive management. Also, it is necessary that our companies get rid of the way of thinking that is limited to the domestic market. Global economy assumes uncertainty. It also provides enormous opportunities for organizations and individuals who are brave enough and

agile enough to adapt. Globalization is the only way for all countries in transition, and there must be insisted on the development of industrial production. Newly industrialized countries of the world have just worked on developing industrial production and offering their industrial products in the global market.

The data collected in the survey were provided by the students of management and business so it was expected that the majority of examinees in this sample wanted to start their own business. The results point at the fact that the situation in the whole Serbia is not stimulating and that attention must be paid to making conditions and encouraging entrepreneurship among young population. All negative factors that hinder the examinees from starting their own business are the results of the absence of an adequate ambience for encouraging entrepreneurship among the young.

Based on the presented results, it can be concluded that future young experts in the field of management are very aware of the fact that the competitiveness of domestic enterprises and the degree of innovation are at a very low level. Creating a business environment and acquiring new technologies represent a strategic framework for establishing the competitiveness of domestic enterprises, according to the opinion of both the basic and the control group.

The economic development and development of the Republic of Serbia requires the development of a competitive knowledge-based economy, new technologies and innovation.

Acknowledgements

This paper was supported by the Provincial Secretariat for Higher Education and Scientific Research, Autonomous Province of Vojvodina, project name: Improving entrepreneurial climate, the analysis of aspects and possible courses of action among youth in the Central Banat region.

References

1. Drucker, P.: Postkapitalističko društvo. Privredni pregled, Beograd, Srbija. (1995)
2. Radojević, Z., Rakonjac, M., Rakonjac, I.: Position, role and importance of manager in production system (in Serbian). In Proceedings of the International Scientific Conference MANAGEMENT 2010. Krusevac, Serbia, 366-370. (2010)
3. World Economic Forum: The Global Competitiveness Report 2016-2017 (2016). [Online]. Available: https://www.weforum.org/reports/the-global-competitiveness-report-2016-2017-1
4. Nikolić, G.: Ekonomija krize. Arhipelag, Beograd, Srbija. (2014)
5. Đorđević, D., Ćoćkalo, D., Bogetić, S.: Novi model upravljanja i razvoj konkurentske sposobnosti domaćih preduzeća. U Zbornik radova Konferencije JUSK ICQ 2011 – Četvrta konferencija Razvoj poslovne izvrsnosti i konkurentnost domaćih preduzeća, Beograd. (2011)
6. Đorđević, D., Ćoćkalo, D., Bogetić, S.: Novi model upravljanja i razvoj konkurentske sposobnosti domaćih preduzeća. U Zborniku radova Konferencije JUSK ICQ, Beograd, Srbija, 71-74. (2011)

7. Bogetić, S., Đorđević, D., Ćoćkalo, D.: Unapređenje preduzetničke kulture mladih u funkciji regionalnog razvoja. Ekonomski vidici, Vol. 18, No. 2-3, 381-391. (2013)
8. Ćoćkalo, D., Đorđević, D., Bogetić, S., Sajfert, D., Minovski, R.: Quality of business, entrepreneurship education and business start up intentions among students in Serbia: Research results. Industrija, Vol. 41, No. 3, 125-145. (2013)
9. Bogetić, S., Đorđević, D., Ćoćkalo, D.: Odnos mladih u Republici Srbiji prema pokretanju sopstvenog biznisa – Regionalni aspekt. Ekonomski vidici, Vol. 19, No. 2-3, 303-312. (2014)
10. Ćoćkalo, D., Đorđević, D., Nikolić, M., Stanisavljev, S., Terek, E.: Development of the entrepreneurial behavior among young people - Research results from the Banat region. In Proceedings of the VII International Symposium Engineering Management and Competitiveness (EMC 2017), Zrenjanin, Serbia17-22. (2017)

Method of Collecting, Processing and Storing Geolocational Data Received From Mobile Devices[1]

M. L. Voskoboinikov, R. K. Fedorov
V.M. Matrosov's institute of system dynamics and control theory SB RAS
Lermontov str., 134, Post Box 292
664033, Irkutsk, Russia
mikev1988@mail.com, fedorov@icc.ru

Abstract. The article considers the method of calculating the average speed of traffic along the roads of the Irkutsk region, Russia. The method implementing software is based on mobile technologies, distributed computing and the use of the principles of open access to spatial data. An Android application and a Web service have been designed for the collection, processing and storage of geolocation data.

Keywords: mobile technologies, distributed computing, GPS

1. Introduction

In the Irkutsk region of Russia, anthropogenic impact on the environment is being investigated. One of the main factors of anthropogenic impact is the transport availability of the environmental objects. To determine the transport availability of the environmental objects, the average speed of traffic along the roads of the Irkutsk region is required to be assessed. At the moment there is no information on the average speed of vehicles at specific time intervals. In the earlier study, the average speed of traffic was assessed based on information on the type of road surface.

The use of public mapping services provided by Yandex, Google and others to calculate the average traffic speed for research purposes is impossible due to the lack of open access to the initial geolocation data, such as the speed of individual vehicles, through their API. This circumstance requires the implementation of a system for collecting, processing and storing geolocation data received from mobile devices. In connection with the foregoing, it is necessary to organize the collection, storage and processing of traffic data on the road network of the Irkutsk region.

One of the handiest ways to get real speed of traffic is to use mobile devices equipped with GPS/GLONASS-sensors, permanent access to Internet and opportunity of install custom applications. This way of collecting data uses a concept reffered to as the phone as a sensor [1]. According to the authors in [2], it is quite correct to talk about the standard means of collecting sensory data using mobile devices. Based on the concept of the phone as a sensor in [3], the authors give examples of systems based on

[1] This work was supported by the Center for Collective Use of the IICN ISEC and the RFBR grant, No. 16-07-00411.

the collection of sensory data using mobile devices, as well as practical implementation of the system to monitor the vehicle's compliance with the prescribed route.

2. Problem definition

To obtain more accurate and complete data, several mobile devices are required to be used at once. In this case, there is a significant increase in the volume of initial data, which leads to the complexity of collecting, storing and processing large amounts of geolocation data.

To solve the above problem, the authors suggest the creation of new methods, software and technology solutions based on:

- mobile technologies;
- distributed computing;
- principles of open access to spatial data and Open Geospatial Consortium (OGC) standards.

3. Method and software

This method is based on the use of several mobile devices to collect geolocation data. The use of mobile devices has several obvious advantages:

1) there is no need to install special expensive equipment with GPS / GLONASS-sensors on vehicles;

2) the majority of modern mobile phones with GPS/GLONASS-sensors has an acceptable accuracy of geolocation for further analysis of geolocation data and the identification of regularities.

To solve the task, the following software systems were developed:

- Android application for collecting geolocation data together with the accuracy, time, speed and altitude information;
- a specialized Web service for processing, storing, analyzing and visualizing collected data.

The initial data for calculating the average speed of traffic on the road network of the Irkutsk region consist of:

1) geolocation data containing the movement speed of the mobile devices;

2) database of roads of the Irkutsk region.

4. Implementation of Android-based applications

To implement the Android-application, the integrated development environment Android Studio was chosen.

The Android-application process execution consists of several stages.

At the first stage, after the application is launched, the Internet availability of connection to Web service is checked. If there is the connection, a request for registration to the Web service is sent to obtain an unique identifier that will be used

later to link the data to a specific user to track the movements on the map of the region, otherwise the application displays an error message and terminates.

At the second stage, a request is sent to the service to obtain an application configuration that specifies the interval for accessing the geolocation service. In accordance with the resulting configuration is performed periodically appeal to the geolocation service for access geo-location data such as latitude, longitude, altitude, speed and accuracy.

At the third stage, the received geolocation data is sent to the Web service in the format of the serialized JSON object via the HTTP protocol. If there is no access to the Web service at the time moment of sending data, the data is saved to the local SQLite database.

The next time of the application start, if the application has an access to the Web service, the data stored in the local database is sent to the Web service wit the same way as above.

5. Implementation of the geolocation data Web-service

In accordance with the main aim, the implemented Web service should correspond to the following requirements:

1) establishing and maintenance of HTTP connections with a variety of mobile devices;
2) distributed storage and processing of a large volume of constantly updated geolocation data;
3) online processing and analysis of the spatial data;
4) the data visualization on an interactive map of Irkutsk region.

When choosing the technologies for implementing the Web service, the following criteria were taken into account:

- openness;
- availability of a complete documentation.
- ensuring stable operation in high load mode.

The software platform Node.js was chosen as the basis for the implementation of the Web-service, as well as the REST architecture of distributed applications. As a database management systems, the object-oriented rational database server PostgreSQL was chosen, which allows efficient processing large amounts of data using spatial indexes R-Tree / GIST.

The table structure “location” is given in Table 1.

Table 1. Table structure “location”

Column name	*Column type*
id	bigint
coord	point
datetime	timestamp with time zone
provider	character varying(32)
accuracy	double precision
Uid	integer
altitude	double precision
speed	double precision
Text	character varying(255)
Rid	integer

To ensure the scalability of the system resources, it was decided to share the load between the three servers. A separate server has been used, which performs two tasks:

1) execution of Web-services;

2) load balancing between multiple servers.

Two servers were used for data storage and processing. The data is distributed randomly between them.

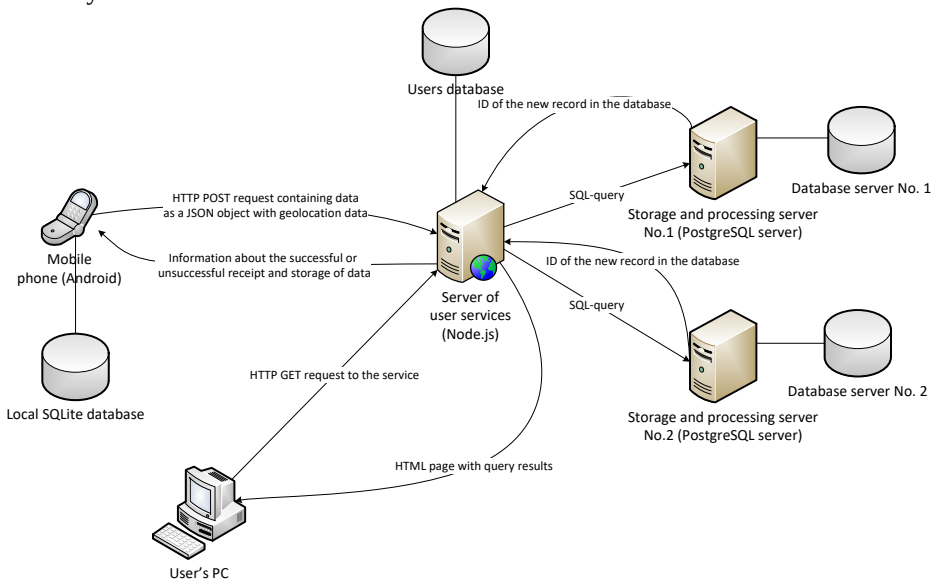


Fig. 1. The system architecture

A PostGIS extension is used to process and analyze geolocation data. For the visualization of geolocation data, an open freely distributed JavaScript library Leaflet was used.

To ensure confidentiality, the Web service does not store any user personal data, such as IMEI of the mobile device, telephone number and other data, which can be used by the attackers in illegal activities.

Fig. 2 is a block diagram of the algorithm for storing the mobile data. The algorithm consists of four steps:

Step 1. The JSON object or JSON object array received from the mobile devices via the HTTP protocol arrives at the system input. The object contains the following data:

— coordinates of the mobile device: latitude, longitude, altitude;
— speed of the mobile device;
— the accuracy of the data;
— data source type, i.e., GPS/GLONASS or mobile network;
— timestamp of the data acquisition taken from the sensors of the mobile device.

Step 2. A SQL query is executed, which uses the coordinates of the mobile device obtained at the previous step and searches for the nearest road in the Irkutsk region road table. The result of the query is the road identifier ID.

Step 3. The obtained ID is saved together with the JSON object data.

Step 4. After successful saving data, a corresponding notification is displayed on the screen of the mobile device.

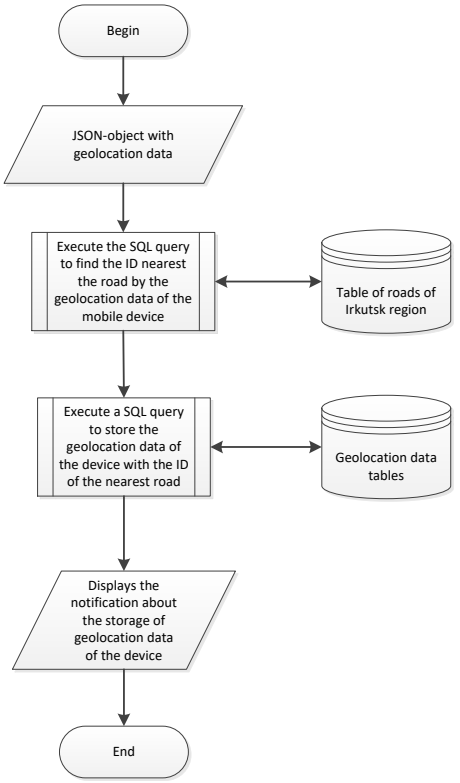


Fig. 2. Block diagram of the algorithm for storing the mobile data

6. Mathematical model

To calculate the average traffic speed in a distributed data storage and processing system, a mathematical model is developed. Below are the symbols of the mathematical model:

K – the number of servers processing data in a distributed system;

L_k – the number of roads on which observations were made on the k-th server;

m_{kl} – the number of stored observations on the speed of vehicles along the L_k-th road on the k-th server;

v_{ikl} – the i-th value of the speed of vehicles on the k-th server on the l-th road;

The arithmetic mean value of the speed of the l-th road on the k-th server can be obtained by the formula:

$$V_{lk} = \frac{\sum_{i=1}^{m_{kl}} v_{ikl}}{m}, k = 1 \dots K, l = 1..L_k \quad (1)$$

The arithmetic mean of the traffic speed for all servers can be obtained from the formula:

$$V_a = \frac{\sum_{k=1}^{K} V_{lk}}{K} \quad (2)$$

Weighted average value of the speed of the l-th road on all servers can be obtained by the formula:

$$V_W = \frac{V_{lk} * m_{kl}}{\sum_{i=1}^{K} m_{kl}}, k = 1 \dots K, l = 1..L_k \quad (3)$$

The weight in this model is m_{kl}. The more stored observations on the speed of vehicles along the L_k-th road on the k-th server, the more weight the k-th server has in calculating the weighted average speed.

7. Implementation of analytical part of the Web service

To calculate the average speed of traffic on the road network of the Irkutsk region, the data presented in Shapefile format were used. It has previously been converted into a PostgreSQL table using the shp2pgsql utility.

The Fig. 3 is a block diagram of the algorithm for calculating and displaying the average speed of traffic on the outskirts of the Irkutsk region. The algorithm consists of five steps:

Step 1. The PNG image of the OSM map of the Irkutsk region is uploaded to the HTML page.

Step 2. oading the vector map of the road network of the Irkutsk region from the set of files format Shapefile.

Step 3. The Web service sends a request to two PostgreSQL servers to calculate the average traffic speed for each road, using the road table of the Irkutsk region.

Step 4. Data received from two PostgreSQL servers is combined to obtain the final average speed of traffic along the roads of the Irkutsk region.

Step 5. Each object of the road loaded on the HTML page is assigned the corresponding value of the average traffic speed, which can be obtained by clicking on it.

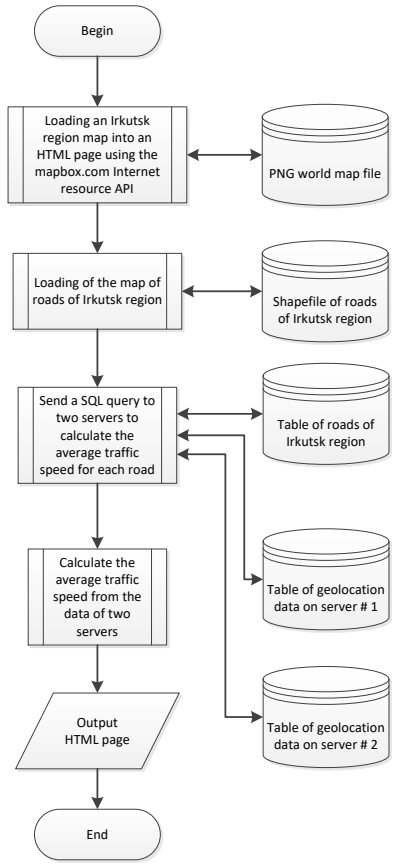


Fig. 3. Block diagram of the algorithm for calculating the average speed of traffic on the outskirts of the Irkutsk region

Conclusion

The result of the work is a system for collecting, processing, storing and visualizing geolocation data received from mobile devices. Data on the average speed of traffic along the roads of the Irkutsk region were obtained. The next stage of the study is to determine the transport availability of the environmental objects of the Baikal region on the basis of the collected information.

References

1. Lane, N.D., Miluzzo, E., Lu, H., et al. (2010). A survey of mobile phone sensing. Communications Magazine, IEEE, Vol. 48, No. 9. P. 140-150.
2. Namiot, D., Manfred, S.-S. *On Open Source Mobile Sensing*. Internet of Things, Smart Spaces, and Next Generation Networks and Systems. Springer International Publishing, 2014. P. 82-94.
3. Neznanov, I.V. Namiot, D.E. Control of transport routes via mobile phones. International Journal of Open Information Technologies. 2015. Vol. 8. P. 30-39 (in Russian)

You Can't Get There from Here: Experiences with a Retraining Course in Web Development

Veljko B. Petrovic[1], Dinu Dragan, and Dušan Gajić

[1] Computing and Control Department, Faculty of Technical Sciences, University of Novi Sad,
Trg Dositeja Obradovića 6,
21 000 Novi Sad, Republic of Serbia,
{pveljko, dinud, dusan.gajic}@uns.ac.rs

Abstract. This paper chronicles experiences in teaching a retraining course in web development using the 'crash course' methodology and presents the challenges of developing such a course under severe constraints of time and subject matter. It analyzes the presented problems under Felder and Silverman's taxonomy of learning styles and introduces the concept of the Minimal Active Knowledge Set (MAKS)—the minimal number of skills required for independent work and adopting active learning styles. It then estimates the MAKS for several contemporary web development technologies, and using this data illustrates the fundamental difficulties of teaching such a crash course. Finally, it presents possible solutions and ameliorations for the fundamental problem so identified.

Keywords: retraining, education, web development, angular

1. Introduction

This paper presents experiences in teaching a retraining course in web front end development as part of a project to increase the number of ICT (Information and Communication Technology) experts in the Republic of Serbia. This project is in keeping with the developmental strategy put into practice in 2010 by the government of Serbia[1] with the express goal of increasing the number of people so educated. Since 2010 the growth in this sector has been explosive with the ICT Cluster of Vojvodina numbering 37 member-companies[2] with over 2500 ICT professionals between them[3] as of 2015. With this growth comes a growing need for ICT experts which the university system with its lead time of 4-5 years simply cannot cope with.

There is, thus, common cause between the companies (who require an expanded workforce to continue their development), the state (which wants to expand the ICT capacities of the country in line with policy goals), and people who wish to find a job in the ICT sector (who want retraining in order to be better able to acquire and retain such a job). This has led to a number of initiatives which seek to enhance the availability of quick ICT education and provide for a smooth expansion of the workforce using retraining. The course described here was held under the auspices of one such project spearheaded by the Ministry of Public Administration and Local Self-Government of the Republic of Serbia[4].

The purpose of the course was to teach the entrants web front-end development as part of a total effort to retrain people without any previous ICT experience into junior-level full-stack developers. To this end, the course was a part of a broader sequence of courses which covered, among other topics, procedural and object-oriented programming, database design and administration, web back-end development, and modern methodologies of software development. Furthermore, the course sought to accomplish the retraining in the field of web front-end development in 20 standard 45 minute classes spread over five four-class intervals over a period of about three weeks. This paper outlines the difficulties this posed and how they may be best tackled.

The paper is organized in four sections. The first is the introduction which introduces the problem and the structure of the paper. The second is the background section which further specifies the problem, offers additional information on education and retraining courses in ICT, provides an overview on the work in the field of learning and teaching styles in engineering education, and introduces the concept of the Minimal Active Knowledge Set (MAKS). The third section uses the MAKS to evaluate web-development technologies. The fourth section concludes the paper.

2. Background

This section outlines the importance of retraining and how it is different from education, the constraints under which the course was forced to operate, previous work in the field, and fourth it uses this work to introduce a new concept, that of the MAKS.

2.1. The Concept of Retraining

Training and, therefore, re-training are not the same thing as education[5], at least in the narrow sense of the terms. Training implies a narrow focus on *how* over *why* and the importance of acquiring immediately applicable skills without necessarily acquiring a broad understanding of some subject. The learning cycle is, thus, shorter, and more focused on the practical aspect. This focus on the practical will become crucial when considered in the light of active learning.

Retraining is even more narrowly defined: not only is it training in the sense described above, but it focuses entirely on providing new job skills for a different career and is entirely focused on results in as short a time as possible. This coincides with the course constraints (subsection 2.2.) which also make a point out of rapid results.

This all means that it is necessary for any retraining course to focus on *how* one solves a certain class of problems and which tools one might employ to this end. Given the paucity of time, the question of *why* certain tools are used or why certain problems are solved in the way they are cannot be given any significant time.

2.2. Constraints of the Course

The course was constraint by its nature in several different ways: it was constrained in **subject,** in **time,** and by the **audience.** Each of the constraints is a natural extension of the purpose of the course, but limits the possible approaches all the same.

Subject constraints. The subject of the course was, naturally, limited to the development of web-based front ends, but further than that, was constrained to provide knowledge that is *current* and *instantly applicable.* The entrants are to be, after the course sequence is complete, accepted as interns in various ITC sector companies, and these companies required that the entrants be capable of full-stack development using modern technologies. Thus, the material of the course must be current and must be such that the entrants can start producing at least marginally useful work immediately.

Time constraints. The entrants don't have a surfeit of time in which to listen to a course, and the companies and the state turned to a retraining course *precisely* because they wanted faster results than what traditional education can provide. This means that the course had to fit into a very limited time slot, specifically, in 20 45-minute classes.

Audience constraints. By the terms of the project the audience was composed exclusively and deliberately of entrants who—while of above-average intelligence—had no previous ITC experience whatsoever. This extended to limited understanding of the operation and function of computers from a user's point of view, and no understanding of the underlying architectures of computer-based systems. This constraint is particularly revealing because most computer science and ITC education and training is deliberately done by people who, on the contrary to the above, *do* have some previous experience and/or have a pre-existing interest in the subject. This means that it is possible that the way courses are conducted has implicit and insufficiently examined previous-knowledge requirements.

2.3. Learning Styles

The question of how engineering students learn has long been studied[6]. According to Felder and Silverman's characterization, the styles of learning/teaching break down according to five dimensions: perception, input, organization, processing, and understanding. This was later modified by Felder[7] by removing the organization pole entirely and renaming the 'auditory' pole to the 'verbal' pole, but the central idea remained and is still the subject of education, speculation, and refinement[8–10] leading to a psychological instrument known as the Index of Learning Styles (ILS)[11, 12].

According to this body of research the way to provide efficient and effective teaching is to teach according to every pole of every dimension. The poles of most dimensions are adequately covered by common teaching methods, but the organization and processing dimensions demand adaptation. This adaptation is a focus on what is known as active learning: learning where the student learns as part of practical work. This method of learning is inherently inductive and active and is highly recommended especially in the light of information from [12] which points out that engineering students tend to be active learniers (as much as 64% of them). The importance of active learning is further bolstered by Armstrong's work on natural learning[13] which finds active learning to be crucial.

When this overwhelming support for active learning strategies is put together with retraining's focus on practical know-how and immediately useful acquired skills it seems that the correct course of action is painfully obvious: the whole course must be built on a firm foundation of scarcely anything more than practical work, independent work by the entrants, and other active learning approaches.

The question is: is this highly desirable state of affairs *possible?*

2.4. The Minimal Active Knowledge Set

The chief problem with active learning/teaching is that it has prerequisites. Specifically, a student must know a minimal number of things before any form of independent work and active knowledge and skill acquisition is possible. For purposes of this paper what a student must know in order to start independent work is termed the Minimal Active Knowledge Set ('MAKS') of a certain technology or subject matter. A great example not related to computer science is the teaching of physics at sub-college level. In order to work independently solving problems in elementary mechanics a knowledge of calculus is required, therefore the MAKS of mechanics includes calculus.

The MAKS is a particular problem for teaching certain developmental architectures in computer science because to accomplish anything several technologies must be mastered independently before they can be used in an interrelated manner to actively learn the developmental architecture in question.

A way to work around MAKS is to, if the technologies being studied are loosely coupled, employ mocking. Much in the way parts of a system being tested may be mocked to enable module/component/unit testing, certain technologies and subjects of study can be abstracted away for later by being replaced by loosely equivalent 'mocked' versions. For instance, a project that's meant to teach graphical user interface development may be rendered independent of knowledge of database stores by wrapping a database in a data access layer which is then provided to the users as a black box.

Mocking is a powerful tool but it is not applicable in the cases of tight coupling between the technologies that make up some sort of developmental architecture. Further, each instance of mocking removes the student from the authentic experience of development and further abstracts the learning environment working, thus, directly on lessening the unique benefits of an active learning/teaching style.

3. Web Development Technologies and Issues

The course constraints, as stated, demanded immediately employable full-stack developers with modern skills. This meant that the only technologies worth discussing are the ones in modern active use. In reality, the one used was Angular 4[14], but for the purposes of this section ReactJS[15], Ember[16], and Vue[17] will also be considered. The key part of the consideration is to merely estimate the MAKS of each, which is to say the number of things that a student must learn before they can reasonably be said to employ active learning strategies—i.e. are capable of

independently solving the simplest possible problems using the technology being discussed.

Methodologically, the way the MAKS was estimated is heavily based on the documentation of the various frameworks, i.e. the estimate was weighted toward what the developers themselves considered to be information one would need to start via tutorials and quick-starts. This helps ameliorate personal bias of the authors, if any, and reduces the subjectivity of the measurement.

Angular. In order for the student to start making simple Angular applications in the interests of active learning they must have first acquired at least basic knowledge of HTML (for providing structure and defining the structure of Angular components), CSS (for the design of the components used and for providing for a reactive interface via style binding), TypeScript (for providing the behavior for the Angular components used), JSON (for configuration), and NodeJS and its configuration and use. Apart from all of this, the student must also learn how Angular works to the extent of being able to write their own components, and employ routing to provide a multi-view one-page application. To be able to do the latter, the concept of annotations and dependency injection must also be covered.

An additional hurdle is that, to write an *actual* Angular application (the preceding allows for toy examples, nothing more) the student must be conversant in asynchronous programming techniques (most notably promises), and further have a REST-capable server, a database, knowledge of the HTTP protocol, of network IO in Angular, and of writing and injecting Angular services. Of all the things listed only the server and the NodeJS configuration can satisfactorily be mocked without impacting the student's ability to learn. Further, these technologies largely can't be acquired piece by piece but must be all present in order for even a simple application to be possible.

React. React is, at first blush, considerably simpler than Angular. The MAKS only requires HMTL, CSS, JSX (a custom templating language used by React), ES6 JavaScript, and a basic understanding of Node.js. However, there is a scope mismatch. React covers only a part of the abilities of Angular. Angular is meant to contain everything that a fully professional web application might need. React merely covers interface rendering. This means that to achieve similar results in an industry-standard way, one needs WebSockets to provide a communication API and Redux to manage state. Now, these frameworks are not *required* and they can be replaced by other frameworks, but this introduces an extra layer of complexity by requiring further Node.js/npm expertise in order to make all of these independent libraries work together without causing complex, recursive issues of code interdependence known under the jargon term of 'dependency hell.'

Thus, the MAKS of React has to be expanded with npm skills and dependency management, knowledge of HTTP and the WebSockets API, and knowledge of immutable structures and functional programming, as well as asynchronous programming should communication with a server be required. On the plus side, React is easier to mock as the various components aren't *meant* to work together as in Angular and, so, it is possible to implement code which is more didactic and less performant, elegant, and so on. This, however, distances the React students would learn from the sort actually used.

Ember. Ember is more like Angular insofar as it tries to provide a full framework. The MAKS for ember involves a specialized CLI command language for generating

elements of the code, the Handlebars templating language, ES6 JavaScript, HTML, CSS, HTTP protocol details, and asynchronous programming with promises which is built into the way Ember handles data models.

Ember does not do well with mocking since everything is built in and most of the above is required to get any sort of application off the ground at all.

Vue. Vue is a framework designed to be highly modular and to expose as little of its API surface as it is possible. The results is that Vue at least *promises* a very manageable MAKS. It even defines it in its documentation: all that is required is HTML, CSS, and ES6 JavaScript. An analysis of Vue's documentation shows that this is nearly true. A prospective user needs to familiarize themselves with a relatively simple templating system (Vue supports but does not require JSX), and to communicate with a server and support one-page multi-view applications a routing module that's built-in must be included. Communicating with a server is not particularly well supported, however, this does mean that it is possible to write the simplest possible asynchronous data handling code which is useful for didactic purposes.

Vue supports mocking best of all: the core is entirely simple, and the rest of the system supports very simple programming without the need for complex frameworks or a huge API surface. Even the server communication can be quite easily entirely mocked by hiding asynchronous data loading behind a pre-route event.

4. Conclusion

This chapter is subdivided into three sections, outlining discussion, suggestions, and results.

4.1. Discussion

Supporting an active learning style from the point of view of the teacher produces four types of problem: prerequisite problems, smoothness problems, setup problems, and rigidity problems. Prerequisite problems revolve around the skills one simply *must* have to use a technology actively. Among the frameworks observed this list is generally uniform: HTML, CSS, either TypeScript or JavaScript, and asynchronous programming and HTTP unless the communication with the server is entirely abstracted. Smoothness problems revolve around moving from one mastery level over a framework to another and how hard it is to resume active learning after a new subsystem is introduced. A smooth learning curve is crucial so that concepts can be introduced one-by-one while maintaining constant active learning. Setup problems are the costs of initial setup: how much dependency and tooling management is required for the student to start working immediately. This class of problem is amenable to mocking as a complete setup can be stored and distributed to the students to use without really understanding how it is built.

Rigidity problems stem from the tendency of web development frameworks to be 'opinionated,' i.e. enforce a certain way of doing things. While such rigidity can be useful in enforcing a uniformity of solutions which improves legibility, simplifies testing, and enhances interoperability, it is not an advantage in teaching/learning as

making mistakes and ‘bad code’ and then improving it is an important learning experience. A framework which only allows production-level code to be written demands to high a jump from the prospective student.

Further research in this filed can hinge on a careful construction of a knowledge pre-requisite tree based on the implementation of a single application in multiple frameworks, and on a more careful quantitative study of the behavior and opinion of students, possibly by tasking them with simple development in a controlled environment and then collecting telemetry as to their behavior and actions.

4.2. Suggestions

Insofar as it is fitting to do so, it is possible to produce some sort of ‘recipe’ for a functional course of this kind. Experiences in teaching Angular in this context have sharply defined those aspects of the Angular stack that course entrants have serious problems with. Students have problems grasping the complex dependency networks that using npm demands, as well as the general shape of the system. Beginning programmers, no matter their intelligence, spend at least a little while at the stage of thinking of one program doing one thing, no more, and have trouble visualizing how the various bits of code (which they individually might be able to understand) and configuration files in various dialects all come together to create on smoothly operating solution. They have trouble finding—in the words of one student—the ‘main’ function to start with.

This leads to two types of tentative precepts regarding course organization: structural and technological. Structurally, a course of this nature must be the last taught following separate courses in pure web design (if needed) and the design of server-side solutions. This way, the subject may be presented as explicitly integrative to the students and explicitly as something which unifies all previous knowledge.

Technologically a non-opinionated framework is necessary. The authors can, based on their experience, recommend either Angular or Vue. Angular is more widely used and its complete integration of nearly all necessary elements of a web application is valuable. However, it is incredibly cumbersome to start using and requires, in the case of minimized MAKS, a great deal of mocking, and promising students that eventually certain things will make sense. There’s only so much trust one can expect of a student in this wise before they feel entirely adrift and give up. Therefore, for people lacking experience, Vue is perhaps the best, and furthermore, for didactic purposes, Vue should be written ‘badly,’ which is to say without using certain popular technologies at first: that means students should be taught simplistic callback-based asynchronous server communication, and should write in the most natural version of JavaScript ignoring, for the time being, such things as transpiling.

4.3. Results

The course had a total of 25 entrants, two of which failed to complete the course for reasons of absence. Of the remaining 23 a further five did not complete the segment which interest us for reasons of absence. This leaves us with 18 students who did, in

fact, complete their assigned tasks. Of those, on a hundred point scale the average score was 62.15 ± 8.76, with a standard deviation of 17.61, and a median of 59.86. The maximum score was 94.75. Of 18 students, 8 earned a course certificate with a recommendation.

Given the necessary pace and scope of the course these are, at minimum, acceptable results.

Acknowledgments

The reported research is partly supported by the Ministry of Education, Science, and Technological Development of the Republic of Serbia, projects TR32044 (2011-2017), ON174026 (2011-2017), III44006 (2011-2017), and III47003 (2011-2017).

References

1. Republic of Serbia: ICT policy. Služ. Glas. 51, (2010).
2. Vojvodina ICT Cluster: Web Presentation of Vojvodina ICT Cluster, http://vojvodinaictcluster.org/, (2017).
3. Milovan Matijević, Milan Šolaja: ICT in Serbia at a Glance. (2015).
4. The Ministry of Public Administration and Local Self-Government of the Republic of Serbia: Добродошли на сајт посвећен преквалификацијама за ИТ (Welcome to the IT retraining site), https://www.itobuke.rs/dobrodosli-na-sajt-posvecen-prekvalifikacijama-za/.
5. Masadeh, M.: Training, Education, Development and Learning: What is the Difference? Eur. Sci. J. ESJ. 8, (2012).
6. Felder, R.M., Silverman, L.K.: Learning and teaching styles in engineering education. Eng. Educ. 78, 674–681 (1988).
7. Shuman, L.J., Atman, C.J., Eschenbach, E.A., Evans, D., Felder, R.M., Imbrie, P., McGourty, J., Miller, R.L., Richards, L.G., Smith, K.A., others: The future of engineering education. In: Frontiers in Education, 2002. FIE 2002. 32nd Annual. p. T4A–T4A. IEEE (2002).
8. Graf, S., Viola, S.R., Leo, T., Kinshuk: In-Depth Analysis of the Felder-Silverman Learning Style Dimensions. J. Res. Technol. Educ. 40, 79–93 (2007).
9. Prince, M.: Does Active Learning Work? A Review of the Research. J. Eng. Educ. 93, 223–231 (2004).
10. Felder, R.M., Brent, R.: Active learning: An introduction. ASQ High. Educ. Brief. 2, 1–5 (2009).
11. Felder, R.M., Soloman, B.A.: Index of learning styles. (1991).
12. Felder, R.M., Spurlin, J.: Applications, reliability and validity of the index of learning styles. Int. J. Eng. Educ. 21, 103–112 (2005).
13. Armstrong, J.S.: Natural Learning in Higher Education. In: Seel, P.D.N.M. (ed.) Encyclopedia of the Sciences of Learning. pp. 2426–2433. Springer US (2012).
14. Google Inc.: Angular. (2017).
15. Facebook Inc.: React. (2017).
16. Tilde Inc.: Ember. (2017).
17. Evan You: VueJS. (2017).

Value Co-Creation using Scrum in Software Development Industry

Savina Čolić[1]

[1] Technical Faculty 'Mihajlo Pupin', Đure Đakovića bb,
23000 Zrenjanin, Serbia
savina.dj@gmail.com

Abstract. Achieving and sustaining competitive advantage in today's marketplace exceed company boundaries, implying active involvement of all participants of the supply chain – outsourcers, suppliers, and end users, in the process. In the situation where the market and its most important participants – consumers – are saturated with existing business strategies, market differentiation becomes more and more difficult to obtain and keep. On the other hand, risk-packed projects prove to be highly doubted for success even by the project members. Due to this situation, companies strive to foster and improve their market position in various ways, minimizing risk and optimizing value from the products they develop. One of the popular methods for doing so is the Agile ones, out of which the most frequently used one is Scrum.

Keywords: agile methodology, value co-creation, risk management

1. Introduction

New world order of strengths combined with growing competition brings upon a paradox reflected in the fact that, even though the consumers have a wide range of products and services available, they are not completely satisfied with the offer; on the other hand, companies invest into greater variety of products, yet they find it increasingly difficult to differentiate from their competition. Therefore, the main concern is how to achieve and retain diversified offer, and subsequently, favorable market position. Another point of concern is sustainable growth, which, combined with the pressure from the competition and other factors, becomes difficult to achieve.

In the situation as presented above, one of the possible solutions for achieving competitiveness is the creation of value for the customer, specifically tailored for their needs. To achieve this in the most effective way, the boundary between the producer/service provider and business customer is removed and the customer is actively involved in the process of product/service development. Therefore, value co-

creation becomes a central strategy for achieving competitiveness and is done with high focus on both sides.

The aim of this paper is to present the implications of mutual work on value creation on B2B level, adding value for the end consumers and bringing various positive effects for the entire value chain.

Being a conceptual work on current development of value co-creation using Scrum framework, this paper is a comprehensive analysis of the literature on the topic, written up to present.

2. Mitigating Traditional Project Management Problems using Agile Methods

Merging work between companies, as presented above, carries certain risk, related to the procedures, ways of doing business, transparency, trust and, increasingly, geographical distribution of the teams. In the IT industry specifically, a large proportion of projects proved to have failed by certain criteria, not managing risk proactively. According to [3], 17% if large IT projects go so badly that they can threaten the very existence of the company. The same source cites 45% budget and 7% time overrun, while delivering 56% less value than predicted. In addition to this, a survey done by Geneca [5] with 600 people closely involved in software development projects showed that around 75% of people doubt the project success even at the start of the project.

On the other hand, as presented in the Introduction, industry competitors try to achieve and maintain their market differentiation and competitive advantage by providing their market with a unique product or service. In order to do so, the modern business world sees an increase in doing work in projects, which, according to PMI [4], are temporary endeavors to create a unique product, service or result.

Thus, the combination of strives to provide unique products and minimize risk leads to the implementation of agile methods in management. Carrying this idea of agility, the work is done in short iterations, result of each being an increment in product development. The focus is on communication, delivering working product (even in early stages of its development – reflected in working functionalities of the product in that case), collaboration and openness. Documentation, strict plans and adhering to standard procedures and tools are therefore less important. In this way, the progress towards a mutual goal is faster than in traditional setups, along with the minimized risk due to the small scopes of work pulled in each iteration of work. Iterations themselves last for anything in the range between one and six weeks.

Agile methodologies and frameworks are numerous. The most frequently used framework is Scrum, due to its scalability and adaptability.

2.1. IT Industry Trends

According to [1], the trends and innovations, including the Internet of Things, already mainstream cloud computing, augmented and virtual reality, drones and autonomous vehicles, will shape the technology industry over the following several years and they came into sharper focus in 2016.

The same source cites current order in this area with a five-strong set of U.S.-based supercompetitors – Google, Amazon, Apple, Facebook and Microsoft. They are closely followed by the 'Next 20', including Adobe, Dell Technologies, Hewlett Packard Enterprise, IBM and others, all shifting from exclusively hardware to software-defined hardware and thus from pure products to services, managed services and solutions. Finally, a group of so-called 'Chinese challengers', including Alibaba, Baidu, Huawei and others, expand quickly and significantly contribute to shaping the competition landscape, both present and future.

The ever changing and increasingly complex situation, as presented, implies a great amount of competition in the software development industry. Thus, it is clear that the pressure posed on market participants is strong, requiring great effort to establish and maintain competitive advantage.

2.2. Delivering Software using Scrum

Just as in other industries, software development companies need to address the problem of competitiveness, risk minimization, as well as the overall process transparency. In addition to the complexity of this situation, faster pace of technology change (compared to other industries) needs to be taken into account as a specific challenge. To mitigate the situation, SD companies mainly use agile methods and frameworks for portfolio/program/project management and product development, among which the most frequently used one is Scrum.

Being a framework applicable to portfolios, programs and projects of any size or complexity, Scrum can be effectively used in any industry to create a product, service, or another result. It implies a collaborative effort of three roles – Product Owner (person representing customer and stakeholders), Scrum Master (person in charge of facilitating Scrum in its full or narrower scope) and the Development team (within which no other roles than Developer is recognized). People carrying these three roles collaborate on creation of a new product, service, or other.

The advantages of using Scrum are as follows:

- Adaptability and flexibility – fast response to changing needs, which are 'by default' in the industry;
- Doing work in iterations – thus minimizing financial, organizational and other risks and enabling realization of benefits early as the product continues to develop;
- Speed – with the first delivery of working features within about a month or less. Early and regular releases result in having higher revenue from incremental delivery;

- Effectiveness – enabled by daily communication, which replaces extensive documentation. The active involvement of a user representative, the high visibility of the product in progress, combined with the aforementioned flexibility to change when change is needed, create much better business engagement and customer satisfaction;
- Transparency – since the provider and the customer are considered a single team. Agile development principles encourage active involvement from the Client's side throughout the process;
- Collective accountability – all participants in the process are equally accountable for the success of the project;
- Continuous progress – the shortness of iterations makes work focused, enabling progress all the time;
- Creation of right product – finally, the combination of all the above mentioned benefits, leads to the ultimate creation of just the right product for the customer.

In addition to these, the cross-functionality and self-organization in teams which are empowered, as well as short, focused work cycles with potentially shippable products resulting from each of them, are the most important advantages of delivering projects using Scrum.

As mentioned in the Introduction, the creation of tailored value is the main goal in today's world of business. With the Client actively involved in the process of product creation daily, the setup makes a good starting point for creation and retaining competitive advantage.

2.3. Scrum Values

Historically, agile software development was introduced by a group of software developing specialists, and based on best practice. The values they put forward were close collaboration with business teams, face-to-face communication, frequent delivery, accepting customers' changing requirements, as well as adaptive organizational capability of teams, according to changing business requirements. These principles stand as more important, over silo mentality, extensive written documentation, delivering complete product and fixed set of requirements, respectively.

Scrum specifically, as one of the versions of agile software development process, has the following five values: commitment, courage, focus, openness and respect. This means that each team member, being part of a self-organizing unit, commits to achieving the team goals; they are courageous to work through challenges (including conflicts) so that they can achieve these goals; only work agreed prior to a sprint is done (ensuring full focus); team members and Client representatives are transparent in communication and finally, team members respect each other's professionalism.

Set up in this way, Scrum enables fast-paced and highly efficient software development. On the human side, it is enjoyable as people are actively involved, cooperating and collaborating instead of being forced to go through comprehensive specifications and write long status reports. Being empowered, the team members enjoy

each other's respect and feel confident to make decisions, ultimately leading to project success.

2.4. Delivering Value using Scrum

It comes as no surprise that the traditional project management approach is soon to be completely abandoned in the software development industry. The pace at which the technology evolves surpasses the time necessary for compiling the elaborate specifications. Therefore, agile project management and Scrum in particular, have changed this situation.

In the majority of cases, even the customer himself is not always sure about what the final product should look like, due to the innumerable amount of options 'on the table'. Therefore, active communication between the product owner, as the representative on the Client's side and the Development team, facilitated by the Scrum master, helps overcome the presented problem and find the optimal solution. Given the aforementioned pace of technology change, the Client's requirements are expected to change during the process and thus the process can be expected to deviate from the initially planned one. Scrum covers this gap as well, following the process closely and adapting to change as soon as the need for change occurs.

In this way, the product value stays as required by the Customer in the given time frame and is able to change if the need for change occurs. This is the main advantage of applying the Scrum framework – aligning the development of product with the current market needs is an essential prerequisite for achieving and retaining competitive advantage in the IT industry market. This value translates to the entire chain, by allowing the customers to enjoy the newest versions of software and updates from them, intermediary company gain and keep their competitive advantage, and stakeholders obtain their revenue.

3. Conclusion

All things mentioned, it is clear that the advantage of agile project management and particularly Scrum is its simplicity. Features are developed and tested in short cycles, providing first results and thus financial benefit even in the early stages of product's lifecycle. Responsibility is shared and the goals are common, with the extensive communiction and a more effective organization of the team, consisting of members from both the Client and the Software development company. The final result is the greater productivity for everyone involved.

Elimination of unnecessary documentation, process, and practice in managing the project makes it possible to work productively and focus on delivering the optimal product, with the value as required by the Customer. This value is up-to-date with the market needs identified by the Client and overflows to all elements of the value chain, each receiving it in the form that fits their needs.

The ever-changing nature of the information technology is expected to remain permanent, thus always seeking for agility; in the future, large multinational projects, including much more people, could be seen done using Scrum (of Scrums) in a larger number than today. Lightweight documentation shall remain in one of its forms, allowing process tracing, but the focus will remain on increasing productivity and thus competitive advantage.

4. References

1. Acker, O., Hagen, H., Hajj, J.: 2017 Technology Trends, PwC, London (2017). [Online]. Available: https://www.strategyand.pwc.com/trend/2017-technology-trends (current July 2017)
2. Bloch, M., Blumberg, S., Laartz, J.: Delivering large-scale IT projects on time, on budget and on value, McKinsey & Company, New York (2012). [Online]. Available: http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/delivering-large-scale-it-projects-on-time-on-budget-and-on-value (current July 2017)
3. Frank C.H., Understanding agile project management methods using Scrum, Purdue University Calumet, Hammond, Indiana (2010). OCLC Systems & Services: International digital library perspectives, Vol. 27 Issue: 1, pp.18-22
4. What is Project Management?, Project Management Institute, Inc. Philadelphia (2017). [Online]. Available: https://www.pmi.org/about/learn-about-pmi/what-is-project-management (current July 2017)
5. Why up to 75% of software projects will fail, Geneca, Illinois (2017). [Online]. Available: https://www.geneca.com/blog/software-project-failure-business-development (current July 2017)

Internet Marketing Strategies for Business Development

Mihalj Bakator[1], Dragica Radosav[1], and Slađana Borić[1]

[1] University of Novi Sad, Technical Faculty "Mihajlo Pupin", Đure Đakovića bb, 23000, Zrenjanin, Republic of Serbia
mihaljbakator@gmail.com, dradosav@tfzr.rs, sladjana.boric@hotmail.com

Abstract. In the modern business world, companies struggle to maintain a stable position on saturated markets. This is due to rising number of new competitors. In order to develop and maintain business, it is necessary to incorporate Internet marketing strategies. These strategies include actions and marketing concepts such as content marketing, SEO, social media marketing and others. In this paper the use of Internet marketing strategies for business development will be addressed. Furthermore, some of the main Internet marketing strategies will be analyzed. The theoretical approach accompanied with analytical research gives and insight of the necessity of Internet marketing for achieving business excellence. Furthermore, the strategies of content marketing, e-mail marketing, and social media marketing will be addressed.

Keywords: Internet marketing, business development, content marketing, SEO

1. Introduction

The Internet is a dynamic technology which opens new development options for businesses. New business relationships are more easily established between companies and consumers [1]. Furthermore, the Internet influenced the growth of international markets through globalization and consumer to consumer relationships.

Internet marketing is a vital tool for business development. It is necessary to address the importance of its applications [1]. Companies can leverage the benefits of Internet marketing in order to achieve better business performance. Furthermore, using Internet based applications and resources increase the marketing capabilities of small and medium enterprises.

The intensity of innovation is increased with the use of Internet for market research [2]. This means that new relationships that occur on social media sites and other online platforms with potential customers increase the business potential of a company [2]. Internet marketing indeed can improve business performance and business development.

Business development is a crucial for the survival of a company on a dynamic market. In addition, the development process incorporates various flexible strategies which are a "fertile" ground for innovative products and services. Furthermore, business performance is dramatically improved [3]. However, there are some aspects of Internet marketing strategies that are not fully examined. There are variables which affect the results of Internet marketing campaigns. It is necessary to take under consideration the

overall positive and negative effects of sometimes intrusive Internet marketing strategies.

In this paper, the application of Internet marketing strategies for business development will be addressed. The paper includes three main sections. The first section explains the concept of Internet marketing and the value it brings to businesses. The second section presents some of the main Internet marketing strategies such as content marketing, search engine optimization, social media marketing and e-mail marketing. The third and final section presents the concept of improving business performance via the use of Internet marketing strategies in small and medium enterprises.

Furthermore, an analytical overview of certain marketing strategies and business development strategies will be given. In addition, the integration of specific Internet marketing models will be presented.

2. Theoretical Approach of Internet Marketing

Researchers define Internet marketing as a dynamic process of customer relationship building. It uses online activities through which innovative ideas are exchanged and new products and services are created [2].

The Internet marketing concept can be divided into five elements [2]:

- First, Internet marketing is a process. The main marketing process includes 4 marketing mix elements or steps. However, internet marketing has seven steps which include creating a defining a business-unit strategy, defining opportunities on the market, projecting and defining the marketing strategy, creating innovative ways to improve customer experience, defining and projecting the marketing plan, creating a user-interface for customers and evaluation of the previous steps.
- Second, Internet marketing includes maintaining customer relationships. One of the main goals of Internet marketing is to find new customers and nurture loyal and committed customers. Furthermore, in order to achieve business excellence, companies have to target three main factors which are important for relationship building with customers. These are awareness, exploration and loyalty.
- Third, the main attribute of the Internet, connectivity and overall presence as an online entity is a main difference from traditional marketing plans and strategies. However, Internet marketing strategies often include offline tools for advertising such as print-media and television advertising.
- The fourth element is exchange. This means that the core of all business is value exchange, more precisely, products and services for financial resources. Furthermore, the relationship between online and offline marketing plans is influential on overall business performance.
- The fifth element is the main goal of both companies and customers. It is the satisfaction of goals and needs which comes as a result of exchange. The evolution of this element nurtures the growth of loyalty of customers towards the company's brand. In addition, satisfaction is necessary for both sides in order to continually improve the relationship with customers and increase business performance.

The concept of Internet marketing is focused on customer relationship management [4]. There are four key stages of relationship management. This is shown on Figure 1. [2].

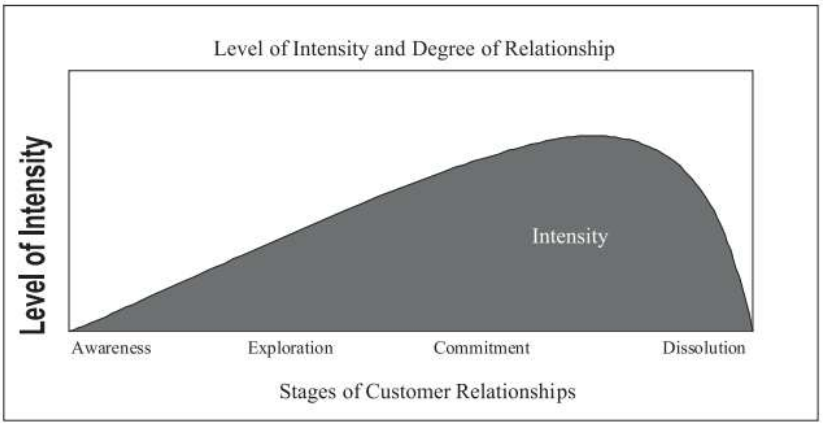


Fig 1. shows the four key stages of customer relationship by level of intensity. The level of intensity of the relationship is the lowest at the stage of awareness. Further, the intensity rises at the exploration stage and peaks between the commitment and dissolution stage. It is clear that online marketing plans engage customers more than traditional advertising plans

3. Internet Marketing Strategies

3.1. Content Marketing in Brand Development

Content marketing is a strong process for developing strong brands. Furthermore, companies which execute online marketing plans and programs, have to focus on creating value in form of various content [5].

When it comes to defining an online marketing plan, content management and defining the strategy behind the content is a necessity in order to create rapport with potential customers [6]. In addition, content marketing strategy development as a part of Internet marketing strategies incorporates two-way communication with the consumers. This ensures further improvement and calibration of the distributed content in the future [5].

One of the main points in content creation is consistent delivery. This ensures a stable growth and helps create brand loyalty. Creating content can be looked at as a form of storytelling where the elements of the story engage existing and new potential customers [7]. However, there are problems which may occur in content marketing execution. This can be intentional, informational content, but sometimes distributing content doesn't bring perspective sales. In order to create sales, it is necessary to engage consumers in such a way where a need is created which has to be satisfied. Engagement is a crucial aspect of content creating and distributing it to customers [6].

The next factor in content marketing is entertaining the audience. This includes creating funny, interesting and overall fun-to-watch content with the goal to target a big

audience. Furthermore, this may include online games and prize competitions where free products or services are given to winners.

The four foundations of an effective content marketing strategy are value, trust, profitable customer action and relationship building [5]. Companies need to develop Internet marketing strategies in such a manner where these four fundamental key elements are included and achieved.

3.2. Search Engine Optimization and Higher Business Performance

Search Engine Optimization (SEO) can be defined as procedures to rank websites higher and increase visibility. These procedures are consisted of operations of crawling, indexing, retrieval and ranking [8]. Crawling can be described as finding content on various pages and websites and retrieving that information in order to define ranks and visibility. Indexing includes registering and allocating websites which are show when a certain phrase is typed in the search engine. Retrieval and ranking is the process of collecting information which the crawlers retrieved and it is used for ranking by the search engine algorithm [9].

Further, SEO includes keyword research, identifying a niche market, code optimization and user experience optimization [9]. These are online activities which are part of the calibration of websites in order to achieve better business performance through higher number of visits by potential customers. In addition, there are offline procedures which include link building and social media usage in order to drive more traffic to specific websites and generate sales [10].

SEO can drastically improve business performance through a stronger presence on the Internet. New customers and potential business partners can more easily find a company that sells specific products or services. However, bad optimization can drive “sterile” traffic which won’t generate sales or improve business performance. This is why it is crucial to develop a well based Internet marketing strategy which includes flexible and well defined actions and procedures.

There are two models when it comes to SEO and advertising through search engines. The first model is Cost-Per-Thousand (CPM) where the company pays a certain, defined amount for every thousand views. The second model Cost-Per-Click (CPC) is based on a fixed cost every time someone clicks on the website of the company who pays for the advertising [8].

Companies have to define their goals, their needs and options before choosing an adequate model for generating views and sales. Business performance is affected hugely by this decision. Projecting, planning a management is necessary to avoid profit loss or bad business decisions.

3.3. Social Media Marketing Strategy

Social media marketing strategies are considered highly successful due to the fact of the sheer amount of people who use social media sites regularly. Researchers have

defined social media marketing as a separate whole from other traditional or even online marketing programs. Surely, it is the pinnacle of Internet marketing [11].

Social media is continuously evolving and changing social interaction between people. Sharing information and data is fast and efficient. New business models emerge from the vast distribution of information and enormous traffic measured in millions of users per day. This dynamic and chaotic environment threatens companies who can't adapt to new rules on the market [12]. Fluctuations on the market are the result of fast information distribution and effective C2C communication. The end-user of products and services defines the market with its opinions posted on social media. Furthermore, the "first-impression" is crucial when it comes distribute new products or services on the market [12].

Companies have to address all the factors which influence profits and product sales. This clearly shows that social media marketing is indeed a "big player" among other online strategies. Consumer habits are dramatically changed due to the fact that product and service research and end-user opinions are forming new market factors.

In the next section, effective e-mail marketing plans will be addressed. The main factors and elements of optimized e-mail campaigns will be analyzed. Furthermore, propositions will be made which include the previous mentioned strategies in order summarize the analyzed aspects of Internet marketing.

3.4. Creating Effective E-mail Marketing Plans

E-mail marketing has changed and today it has a solid impact on business performance. Sending e-mails to somewhat strangers can be risky and often gives poor results. This is due to the fact that there are many individuals, groups, organizations and companies overall practice e-mail marketing in various forms [13].

In order to achieve business excellence and gain positive feedback, it is important to develop effective, non-intrusive e-mail marketing plans. There are three main phases of modern e-mail marketing plans [10].

The first phase is getting permission from potential customers. This is done through initial, positive, short, approval-waiting contact e-mail. Furthermore, after the first positive feedback, there are various types of e-mail actions. Some of the most successful ones are free downloads of useful data or information, free e-books and update lists.

The second phase is accumulating a solid number of subscribers to an e-mail. In this phase, companies have to "listen" to their potential customers and manage their expectations, needs and wishes. In addition, a newsletter can be added to the marketing actions. Auto responders are not welcome. However, a concisely written automatic e-mail may have a positive effect on the consumer.

The third phase includes segmentation and analytics. Segmentation is regulated through filtering some of interests of potential consumers and send offers which may satisfy their needs. Analytics include tools which show and present data of engagement, website visits and response times. Furthermore, the marketing plan has to be optimized in a way that ensures good business performance.

Based on the previous sections and the concept of e-mail marketing the following propositions are defined:

Proposition 1: Internet marketing strategies ensure good business performance through well managed customer relationship management.

Proposition 2: Social media marketing is an inevitable tool for achieving good relationship with potential customers a build brand loyalty.

4. Improving Business Performance

Business performance can be presented as the ratio of sold products or services in a specific period of time. The main factors are profits margins, frequency of sales and customer satisfaction [14]. After the analysis of the benefits of Internet marketing it can be observed that with content marketing, SEO, social media marketing and e-mail marketing, companies can create stronger main marketing plans and acquire better insight into the potential market.

Business performance is improved through better publicity, higher number of sales, satisfied and loyal customers. Furthermore, through a stable and continuous Internet campaign companies can stabilize their business actions on the market. It can be suspected that presence on the Internet is an imperative for a company's survival. The analytical view of the improvement on business performance is given on Figure 2.

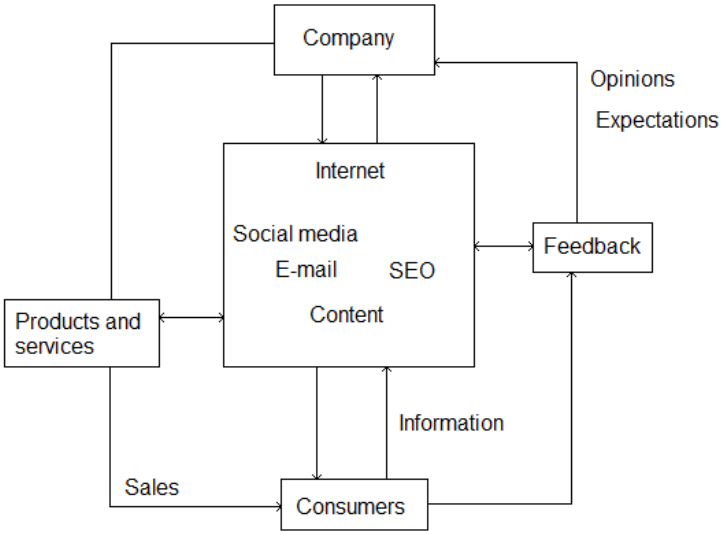


Fig 2. shows the information distribution from the company to the consumer. It is shown, that the feedback element can be direct or go through the communication canals on the Internet. Sales are generated after a certain set of actions and communication between the company and potential buyers

It is necessary to mention that business development in any intensity requires strategic planning and often includes marketing planning. Although, it is not always crucial to incorporate Internet marketing into the development process, it surely has a big impact on the outcome of new business strategies. This is why it is necessary to

optimize the goals and thickly plan every action. The modern business world created dynamic markets where companies can choose: adapt and survive or perish.

5. Conclusion

It can be concluded that Internet marketing strategies can increasingly support business improvement and development. Content marketing, search engine optimization, social media marketing and e-mail marketing are important tools for increasing business performance. There is an imperative for companies to develop highly effective marketing strategies in order to achieve higher profits.

There are issues with this research which have to be corrected in future work. These issues include the lack of external factor analysis and poor addressing to various market segments where somewhat different elements and competitors are present.

For further research it is recommended to conduct case studies with various Internet marketing strategies. Furthermore, the results should be thoroughly analyzed. In addition, external factors should be taken into consideration. This research lacks empirical results, and merely focuses on the theoretical and analytical approach. The goal of this research was to create a starting point from which further research can be conducted.

More studies are necessary to explore the complex problem of the efficiency of Internet marketing strategies for business development. It can be concluded that the approach taken in this paper, showed that there could be a positive correlation between the intensity of Internet marketing strategies and improvement of a company's business performance. Other complementary studies and extensive research is necessary to prove the conclusions which are presented in this paper.

References

1. Mathews, Shane, Constanza Bianchi, Keith J. Perks, Marilyn Healy, Rumintha Wickramasekera: Internet marketing capabilities and international market growth. International Business Review, Vol. 25, No. 4, 820-830. (2016)
2. Sharma, Nandini, Mallika Ahuja, Richa Sharma: Models of Internet Marketing. International Journal of Academic Research & Development JAR&D, 51. (2016)
3. Smyth, Hedley: Market management and project business development. Routledge, (2014)
4. Lind, Mary R., Evetta Culler: Information technology project performance: The impact of critical success factors. In Perspectives and Techniques for Improving Information Technology Project Management, IGI Global, 39-50. (2013)
5. Claesson, Amanda, and Albin Jonsson: The Confusion of Content Marketing: a study to clarify the key dimensions of content marketing. (2017)
6. Einstein, Mara. Black Ops Advertising: Native Ads, Content Marketing and the Covert World of the Digital Sell. OR Books. (2016)
7. Ahmad, Nur Syakirah, Rosidah Musa, Mior Harris Mior Harun: The impact of social media content marketing (SMCM) towards brand health. Procedia Economics and Finance, Vol. 37, No.1, 331-336. (2016)

8. Vo, Tuan: Search Engine Optimization and Its Importance for Business Visibility and Branding. (2016)
9. Srivastava, Shashank Narayan, Suhani Kshatriya, Rajkumar Singh Rathore: Search Engine Optimization in E-Commerce Sites. International Research Journal of Engineering and Technology, Vol. 5, No. 4. (2017)
10. Fahlström, Kamilla, Caroline Jensen: Search Engine Marketing in SMEs: The motivations behind using search engine marketing. (2016)
11. Keegan, Brendan James, Brendan James Keegan, Jennifer Rowley, Jennifer Rowley: Evaluation and decision making in social media marketing. Management Decision, Vol. 55, No. 1, 15-31. (2017)
12. Alves, Helena, Cristina Fernandes, Mário Raposo: Social Media Marketing: A Literature Review and Implications. Psychology & Marketing, Vol. 33, No. 12, 1029-1038. (2016)
13. Hartemo, Mari, Mari Hartemo: Email marketing in the era of the empowered consumer. Journal of Research in Interactive Marketing, Vol. 10, No. 3, 212-230. (2016)
14. Wirtz, B., Pistoia, A., Ullrich, S., Göttel,B.: Business models: Origin, development and future research perspectives. Long Range Planning, Vol. 49, No.1, 36-54. (2016)

The Use of Neural Networks for Customer Experience, Promotion, Brand and Quality Research

Mihalj Bakator[1], Dragica Radosav[1], and Nikola Petrović[1]

[1] University of Novi Sad, Technical Faculty "Mihajlo Pupin", Đure Đakovića bb, 23000, Zrenjanin, Republic of Serbia
mihaljbakator@gmail.com, dradosav@tfzr.rs, petrovic.n26@gmail.com

Abstract. Neural networks have a wide range of application. In this paper the use of these neural networks for market research will be presented. This includes data processing from surveys which incorporate variables such as customer experience, promotion, brand, and product quality. The data was collected via survey, and processed in Microsoft Excel. Furthermore, the processed data was inserted into Matlab. Subsequently, a neural network was created using Matlab. The aim of the paper is to show how precise can a neural network be when it comes to predicting outputs based on inputs. Furthermore, the network was trained with the mentioned collected data. The results are interesting and give room for further research in this field. In addition, it shows how neural networks can help in solving hypothetical problems.

Keywords: Neural networks, customer experience, promotional activities, brand, quality

1. Introduction

In the modern business world, customer experience is an important factor that influences business performance [1]. Companies have to create value for consumers in order to achieve a competitive market position. Creating value for consumers helps building strong brands. Customer loyalty is built with promotional activities and customer satisfaction by product quality [2]. Customer experience starts developing from the moment when the potential customer has first contact with a product or service till the moment he or she stops using the product or service [3].

Brand development and management is an important strategic function that increases business efficiency on a specific market segment [1]. Furthermore, developing a high quality product is an imperative for creating a strong brand. As a result, customer trust increases further resulting re-purchasing a product or service [3].

The research was conducted via survey. The survey results were analyzed in Microsoft Excel and were further processed with the Matlab software. A neural network is created with the survey results in order to train the network so new predictions can be made regarding customer experience, brand and product quality and promotional activities.

The goal of this research is to prove or dismiss the main hypothesis based on the analyzed and processed data. The number of surveys filled by mobile phone users is 782 (N=782). The main hypothesis is:

H_0= Increasing the intensity of promotional activities, product quality and brand strength, increases customer satisfaction.

The alternative hypothesis is:
H_A= Increasing the intensity of promotional activities, product quality and brand strength, doesn't affect customer satisfaction.

The main, general hypothesis can be proved by testing and proving testable hypothesis. The testable hypotheses are:
H_1: Strong advertising increases the popularity of the brand.
H_2: Increasing the intensity of advertising, increases the perceived product quality of mobile phones.
H_3: The stronger the brand, the better the perceived product quality.
H_4: The better the product quality, the better the customer experience.

The paper consists of three main sections. The first section is a short theoretical approach of neural networks, brand, product quality, customer satisfaction and promotional activities. The second section explains the methodology of creating and using MATLAB for creating a neural network. The third section shows the results and addresses the testable, null and alternative hypotheses.

2. Theoretical Approach

2.1. Neural Networks

Neural networks can be presented as a simplified model of the human brain, and they simulate its biological process of learning. Furthermore, neural networks contain a big number of simple, neuron-like elements. Knowledge in the neural network is stored in the main structure of the network. This includes the number of levels in the network, number of knots in the levels and the type of the connections [4].

The "intelligence" of neural networks is located in the collective behavior of the neurons, where each of them does a very limited operation. Neural networks have a wide range of application. They can be used in managing solar power systems [14]. Further, artificial neural networks can find application in reliability analysis that use back-propagation or radial basis functions network [15]. Similarly, neural networks can be used for simulation, optimization and performance prediction [16]. In the future, their use will expand to other sectors in various scientific and social fields [5]. Typical characteristics of neural networks are: adaptive learning; self-organization; error tolerance; operations in real-time and parallel information processing [4]. Neural networks "learn" through data processing. The network finds characteristics and

patterns in data and based on these inputs the network creates “weights” and “learn” new information. Adding new data, the neural network calibrates its “weights” in accordance of the learning algorithm. Decisions are made on the basis of the “weights” that are present in the neural network [5]. Neural network weights

Neural networks can be trained in two different ways. First, the network learns information by monitoring data and adjusting accordingly. Second, learning without monitoring is where the network “learns” by self-modifying the “weights” of the neurons [5]. The advantages of neural networks are the options of wide troubleshooting in various fields. In addition, networks are used to troubleshoot smaller structural problems. However, the limitations of neural networks are big data processing with arithmetical functions. Further, neural networks don’t offer explanation of how they got the solutions. In addition, a large set of data is needed to create a reliable network [6].

In a different manner, neural networks can be viewed as a whole which consists of many connected processors. These processors are called neurons. Each of these processors produce a sequence of real-valued activations. Furthermore, input neurons are activated via sensors that detect external actions and used it as stimulus. The other stimulus is defined by weights [13].

2.2. Promotion, Brand and Product Quality

Promotion can be defined as a communication process with buyers and consumers with goal to increase the number of product and service sales or to improve the market position on a specific market segment [7]. There are five main promotional activities []. These are advertising, sales promotion, public relations, personal sales and direct marketing.

Product brands “help” consumers to decide which product to buy. Often, the brand defines if some product or service will be bought or some other product or service with the same use will be chosen by the consumer [1]. Brands can be viewed as a permanent equity of a company. In order to achieve business excellence, companies have to carefully develop and manage brands. Furthermore, the value of a brand and its power to develop loyalty with the consumers is crucial for business development.

Product quality can be defined as an integrated whole of attributes, design, functionality which can satisfy the needs of consumers [8]. Quality that is experienced by the consumers, or to be more precise, the users of products and services is defined as perceived quality [9]. Companies are tending to synchronize the objective quality and perceived quality. This process is made with innovations and modifications which are based on user experience and feedback. Integrated marketing communication is presented as integrated communication channels through which the company communicates a clear and unequivocal message with the goal to promote itself and its products and services [10]. This type of promotional activity suggests the connection of all segments where the company meets the consumer. In the next section customer experience will be addressed.

2.3. Customer Experience

Companies collect, analyze and use information gathered from consumer with the goal to create precisely designed marketing strategies and to create long-term relationships with the consumers [1]. Furthermore, companies through the loyalty of consumers acquire a perspective of what attracts and keeps customers. Brand development has a key role in customer satisfaction [11].

Consumer trust can be defined as the belief of the consumer that the company focuses on actions from which the consumers will benefit. Trust between a company and consumer contributes to developing a strong brand [10].

Creating value for consumers includes integration of positive elements which are promised by the company. Furthermore, when a company presents its products and services it doesn't promise only one main characteristic, rather a whole congregation of elements is presented in order to give more value than other products on the same specific market segment [1]. Consumers buy products to satisfy their own needs and wishes. Companies develop products and services with the goal to satisfy the needs of consumers [10]. Satisfaction of consumers can be achieved with adequate product quality and functionality. Customer experience is a strong factor that defines the future of a company. Consumers are influenced by the perceived quality and service quality and this affects the image of the company [9].

3. Developing a Neural Network

The collected data was processed in Microsoft Excel. Further, the data was inserted into Matlab and based on inserted data a neural network was created with the *nntool* function. Matlab is widely used for neural network engineering. The Matlab Neural Networks Toolbox offers robustness, completeness, precision and reliability. In addition, Matlab is regularly updated. Furthermore, Matlab allows integration with other applications through several application program interfaces (APIs) such as C and Fortran [12]. The neural network consisted of 10 neurons a two hidden layers. As input the data from brand, product quality and promotional activities were taken. As output, customer experience was inserted. With these settings, the network "learned" and trained. On Figure 1. the neural network training tool *nntraintool* is shown after the training session of the created neural network. On Figure 2. the "Training state" diagram is shown.

Learning rate of neurons can be described as the speed of how quickly does a network abandon old, information/data/"knowledge" in order to accept new information or "belief" [17].

Epoch can be described as a single finite episode/cycle of information processing and spreading of neuron activation. In other words, epoch show the number of times the neural network algorithm sees the entire set of data [13] .

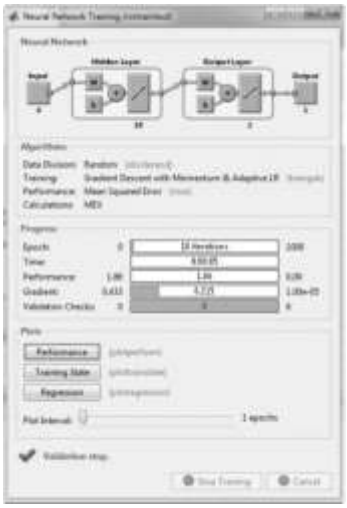


Fig 1. shows the window of the nntraintool in the Matlab software. The progress bars present the epoch number, time, performance, gradient and validation checks

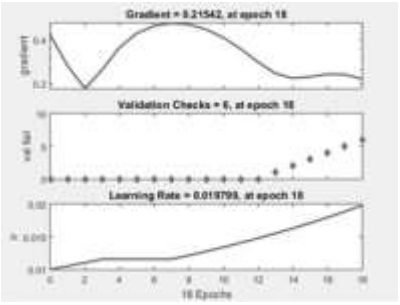


Fig 2. shows the window the training state diagram. It includes gradient, validation checks and learning rate

On Figure 2. it can be seen that the gradient is 0.21542 at epoch 18, the number of validation checks is 6 also at epoch 18 and similarly the learning rate is 0.019799 at epoch 18. The inputs given at the beginning of developing a neural network were medians of the answers on surveys. The output was also median data, and it was used for training the network. In the next section the results of the research will be given.

4. Results

The whole purpose of developing a neural network was to try and predict outputs based on inputs. For this research, the network is trained in a way that it can "predict" with certain accuracy if a testable hypothesis is accepted or dismissed. Figure 3. shows the regression plots of training, validation, test and overall regression.

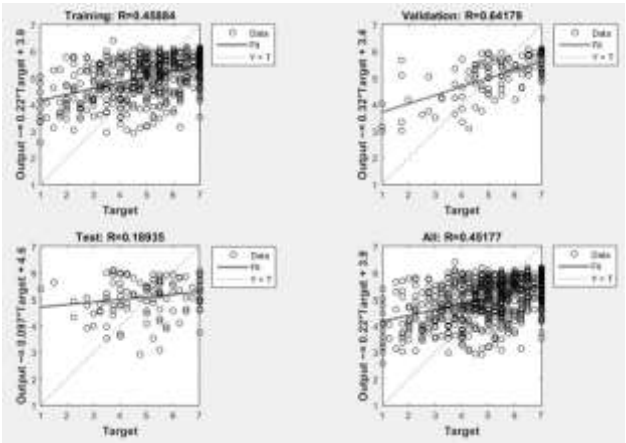


Fig 3. shows the regression plots of training, validation, test and overall regression. Even though regression under 50% is not good, the subjects of research were mobile phone users, whereas their opinions were measured. Regression under 50% is acceptable if the factors include human behavior or human opinions

Furthermore, after the neural network training is complete, the network was used to simulate predictions based on new inputs. For every entered value for an input, the output matched the pattern that was present in the researched data. It can be said that the testable hypotheses can't be dismissed due to the fact that the input and output correlations were matching the testable hypotheses values and variables. In addition, two correlational scatter diagrams were created with the goal to easily visualize the correlation between product quality, promotional activities and customer satisfaction. On Figure 4. the correlation between product quality and promotional activities is shown. On Figure 5. the correlation of customer satisfaction and product quality is presented.

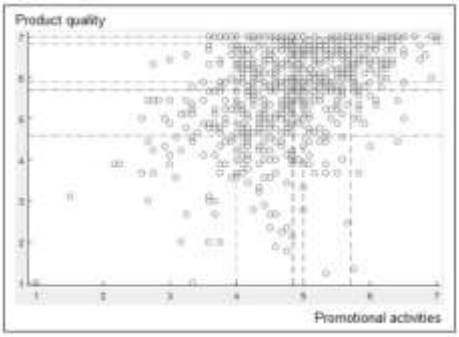


Fig. 4. correlation between product quality and promotional activities. The diagram clearly shows the high concentration of answers in a swarm like shape. This indicates that there is a positive correlation between product quality and promotional activities

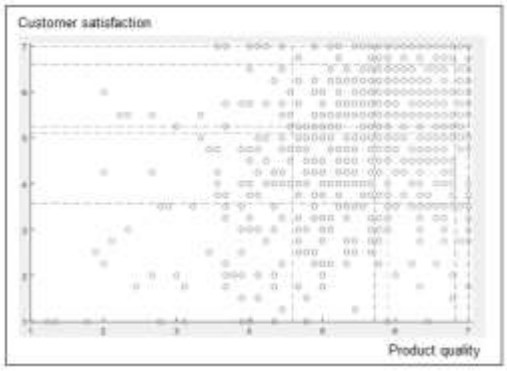


Fig 5. correlation between customer satisfaction and product quality. An arrow shape of the dots present high positive correlation between customer satisfaction and product quality

Based on the research results and presented diagrams, it can be seen a positive correlation between the researched variables. Correlation doesn't mean always causation. However, the theoretical approach and research provides a strong basis on which these assumptions are defined. The research can be finished with the following proposition: *Proposition: Increasing promotional activities and developing high quality products, the intensity of customer loyalty towards the company increases.*

5. Conclusion

Knowledge based systems can have a wide range of applications. Neural networks can be used for predicting outputs based on inputs. This can be used for proving or dismissing testable hypotheses. Based on this research, the developed neural network gave results which clearly showed that the general hypothesis H_0*= Increasing the intensity of promotional activities, product quality and brand strength, increases customer satisfaction*. can't be dismissed.

Creating a neural network in the Matlab software, gave an insight into new ways of research and solving hypothetical problems. The results gained from data processing in Matlab are favorable and acceptable. A positive correlation was determined between the researched variables. Furthermore, these results show that the testable hypotheses can't be dismissed. The learning rate at 0.0197 is adequate considering the average learning rates vary from 0.1 to 0.00001. The regression rate at 45% is satisfying as for human behavior analysis a regression below 50% is acceptable. Implication of this paper include the use of neural networks for predicting consumer wants and needs. Practical use can be found in marketing environment prediction.

The research conducted in this paper has its limitations. The Matlab software gives an insight of hoe artificial neural netwroks can be used for specific prediction and simulation. In contrast, the limitations of hardware performance can limit the siye and complexity of a neural network. However, this research gives a solid foundation for further data collectiong and analyzing. It is recommended to optimiye the number of questions in the suerveys.

Future research has to be conducted in order to solidify the proposed conclusions. Based on those future research results, it can be defined how new brands emerge and penetrate seemingly saturated markets. The use of neural networks for simualtion and prediciton purposes, opens new doors for advanced hypothetical research.

References

1. Kotler, P., Keller, L. K.: Marketing management. Englewood Cliffs, NJ: Prentice Hall. (2016)
2. Wu, C.: Matching Value and Market Design in Online Advertising Networks: An Empirical Analysis. Marketing Science, Vol. 34, No. 6, 906-921. (2015)
3. Keller, K. L., Lehmann, D. R.: Brands and Branding: Research Findings and Future Priorities. Marketing Science, Vol. 25, No. 6, 740-759. (2006)
4. Hotomski, P.: Sistemi veštačke inteligencije. Zrenjanin: Tehnički fakultet "Mihajlo Pupin". (2006)
5. Kendal, S. L., Creen, M.: An introduction to knowledge engineering. Springer London. (2007)
6. Wang, Tong, Huijun Gao, Jianbin Qiu.:A combined adaptive neural network and nonlinear model predictive control for multirate networked industrial process control. IEEE Transactions on Neural Networks and Learning Systems, Vol. 27, No. 2, 416-425. (2016)
7. Đorđević, D., Ćoćkalo, D.: Osnove marketinga. Zrenjanin: Tehnički fakultet "Mihajlo Pupin". (2010)
8. Đorđević, D., & Ćoćkalo, D.: Upravljanje kvalitetom. Zrenjanin: Tehnički fakultet "Mihajlo Pupin". (2007)
9. Percy, L.: Strategic integrated marketing communications. Routledge. (2014)
10. Kotler, P., Armstrong, G.: Principles of Marketing. Essex, England: Pearson Education Limited. (2014)
11. Fraering, M., & S. Minor, M. Beyond loyalty: customer satisfaction, loyalty, and fortitude. Journal of Services Marketing, Vol. 27, No. 4, 334-344. (2013)12.
12. Mendez, J., A., Rosello, G., E., Lado, J., M., Dacosta, G., J. & Cota, M., P: Integrating Matlab Neural Networks Toolbox functionality in a fully reusable software component library. Neural Computing & Application, Vol. 16, 471-479. (2007)
13. Schmidhuber, J.: Deep learning in neural networks: An overview. Neural networks, 61, 85-117. (2015)
14. Karabacak, K., & Cetin, N.: Artificial neural networks for controlling wind–PV power systems: A review. Renewable and Sustainable Energy Reviews, 29, 804-827. (2014)
15. Chojaczyk, A. A., Teixeira, A. P., Neves, L. C., Cardoso, J. B., & Soares, C. G.: Review and application of artificial neural networks models in reliability analysis of steel structures. Structural Safety, 52, 78-89. (2015)
16. Mohanraj, M., Jayaraj, S., & Muraleedharan, C.: Applications of artificial neural networks for thermal analysis of heat exchangers–a review. International Journal of Thermal Sciences, 90, 150-172. (2015)
17. Graupe, D.: Principles of artificial neural networks World Scientific. Vol. 7). (2013)

Management of business processes

Sanja Stankov[1] Slađana Borić[1] Dragiša Tolmač[1] Aleksandar Rakić[2]
[1] Technical Faculty "Mihajlo Pupin" Zrenjanin University of Novi Sad, Djure Djakovica bb
23000 Zrenjanin, Republic of Serbia
{sanja.stankov}sssanja89@gmail.com

[2] Faculty of applied ecology "Futura", Belgrade, University Singidunum,Pozeska 83
11000 Belgrade, Republic of Serbia
{alexandar.rakic}rakic89@gmail.com

Abstract. Since the mid - 90's, the business process management - BPM (Business Process Management) is one of the most commonly argued topics. The reason for this is not a trend, which is very often the case, but the very nature of BPM that returns the focus on what is important in any business - the job! Actually, managing business processes or not - it is quite an unnecessary question. Managing the company, managing the business means nothing more but - to manage a business processes. Even the question: Do you manage your business processes? - the question that is often opened by consultants for their discussions with customers – is just as superfluous. Of course not every company manages its business processes - the only issue that can offer different answers is - in any way.

Keywords: BPM, Business Process Management, Managing Business Processes

1. Introduction

Business Process Management (BPM) is a systematic approach to improving business, based on the design, measurement, analysis, improvement and process management.

Since people are placed inside the rectangle representing the departments in the organizational scheme, it often seems that the edges of these rectangles become stiff barriers. Communication through these barriers is limited, members of the department perform only those tasks that are customary for their department. Each department seeks to maximize its influence and powers, while at the same time optimizing the level of performance of its own department alone. Usually the result is conflicting goals and conflicting actions between different departments and the result of the company as a whole worse than expected.

The business process would then include all the activities that are between the starting point and the endpoint. An example of one of the activities is to notify the warehouse manager that it is necessary to prepare and send a certain number of

products for the shipment. The warehouse manager then has several activities that really do. All these activities are part of the business process.

Individual business processes are in practice rather unpredictable, undocumented and unadjusted by real needs. Dealing with different functional organizational units, process data is often recorded in different software, whereby the problem is a transition from one organizational unit to another or a transition from one to another software. Key business processes that influence the strategic success of an organization include several business functions, and they are the drivers of change from vertical to horizontal forms of organization in modern business.

The problem that is often recognized in the effort to look at all aspects of business management is - the lack of a "common language" at different levels of the company. Very often, the business sector and the IT sector do not understand each other. They “speak” on the same subject, but in a different language. Very often this misunderstanding leads to compromise solutions that do not meet any side - not the business, not the IT.

2. BPM – a project or a process?

Selecting a single platform for business process management of a company is not an easy task. However, setting up the basic question will greatly ease the task - what do we actually want to do? Plan and successfully complete the project? Buy something of good quality at the best possible price? The quality of the decisions is difficult to assess. At the moment when the decision is made, it looks like the best one, but later it may turn out that there were better alternatives [3]. When such a decision is made, we should take a step back. How do we actually see the BPM? What is the BPM for us? To develop and establish a system of BPM can be a project. But the BPM system itself is certainly not and should not be viewed as a project! The reason for that is simple - the BPM is a process, the BPM must be an ongoing activity. This constant activity, something that will have a major impact on the future work and the result of the company must have some methodological basis.

3. What is the life cycle of a BPM?

A platform that responds to such requests and is usually hoisted by independent research companies as the best tool for modeling and analysis of business processes is - ARIS. Why ARIS? ARIS is a tool that uses the methodology of describing business processes. Thus, the tool is designed in a way to allow the inclusion of all elements essential for the business process in a business model itself - the organizational structure, products / services, functions, business events, IT sectors, data, links between the objects and so on. Besides, ARIS Platform answers the question of BPM’s life cycle.

In fact, some tools available through the ARIS platform support each phase of the life cycle of BPM in line, of course, with the methodology and the way of BPM’s life cycle. The platform powerful enough to support you through all stages of the life cycle

of BPM - that is, from strategic planning of business processes, through their design and implementation to the phase controlling – is the only real choice for those who want to turn the business processes and the way they are managed, to their competitive advantage. What makes ARIS different is a three-tiered architecture that enables the repository of business processes to enable the involvement of all levels of the company, but also to support all of the above by its development, the phases of BPM's life-system.

Sometimes companies search for tools that will help them, sometimes in a bit trendy particular, initiatives related to the currently topical issues - eg. SOA (Service Oriented Architecture), EA (Enterprise Architecture), BPA (Business Process Analysis), Performance Management, Risk Management, Compliance, etc. Already today, the business processes are "meandering" through each of the topics listed. Not only is it quite clear that each of these topics in some way "refers" to the topic of BPM, but it is possible to conclude that it was the design of business processes that is necessary for a quality approach to each theme.

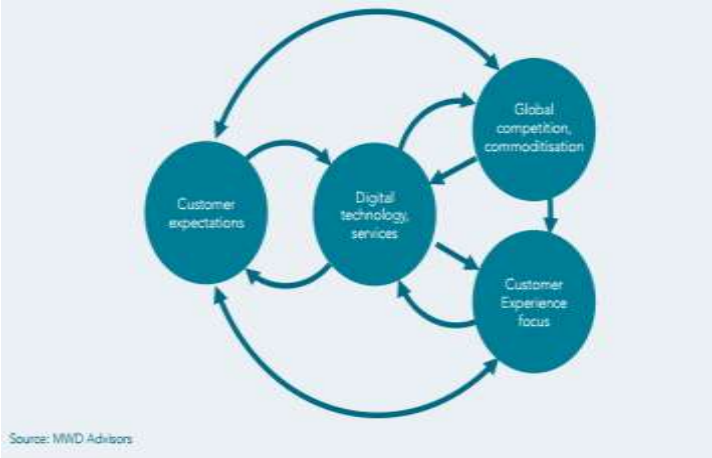
4. Software AG's ARIS Business Process Analysis Platform

The software AG's ARIS Business Process Analysis Platform is ideal for the analysis and the optimization of business processes for better productivity. The journey includes all the interactions the customer has with your organization, from the rise of digital commerce, customer expectations it has spread further and faster than ever before.

The value in mapping helps your organization to see itself in a new light and to build momentum for positive interaction with customers. A single process focuses on departments across the organization to work together in a way like never before, and on providing practical, independent industry insight in business analytics, process improvement and digital collaboration of experts who work with the help of technology.

The MVD advisers are specialists in an advisory firm that provides practical, independent insights in the industrial business, analytics, and process improvement with the help of technology. The access combines a flexible, pragmatic mentoring and advisory services, built on best practices and technologies.

Figure 1. How digital technology and service are the originators of changes and challenges



- This is the process through which a consumer comes into experience with your organization through four main stages:

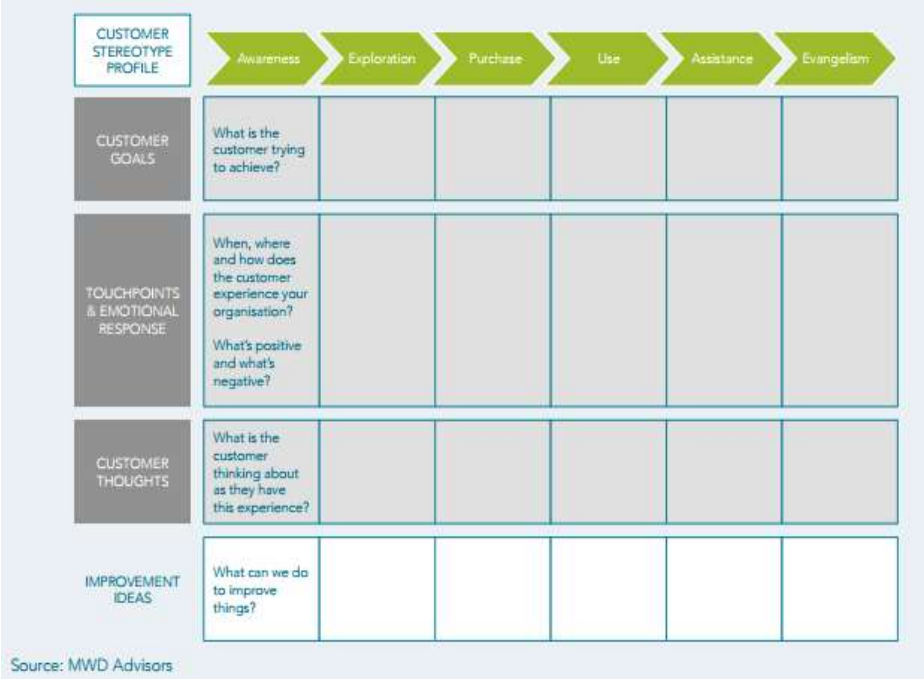
- Awareness; the start of "journey" when the customer comes to us.

- The research; in most cases, the potential buyer will then begin to explore alternative ways to satisfy its needs, where they can explore more organizations.

- The shopping: assuming that the need is still present, the customer decides to buy a particular product or service that meets its needs.

- The use: the customer uses the product or service that is bought.

Figure 2. The example of a map of customer's "journey"



Maps and mapping yield value in two ways, it is important to understand that the act of mapping of the customer's journey gives value in at least two different ways. The first type of value is derived from the results of the mapping: Customers' journey provides valuable tools that encourage your organization to be seen in a new light and to build momentum for projects that will make positive changes in customer needs, while the second type of value comes from the process of mapping their journeys.

Figure 3. The example of customer for a trader: BOPIS

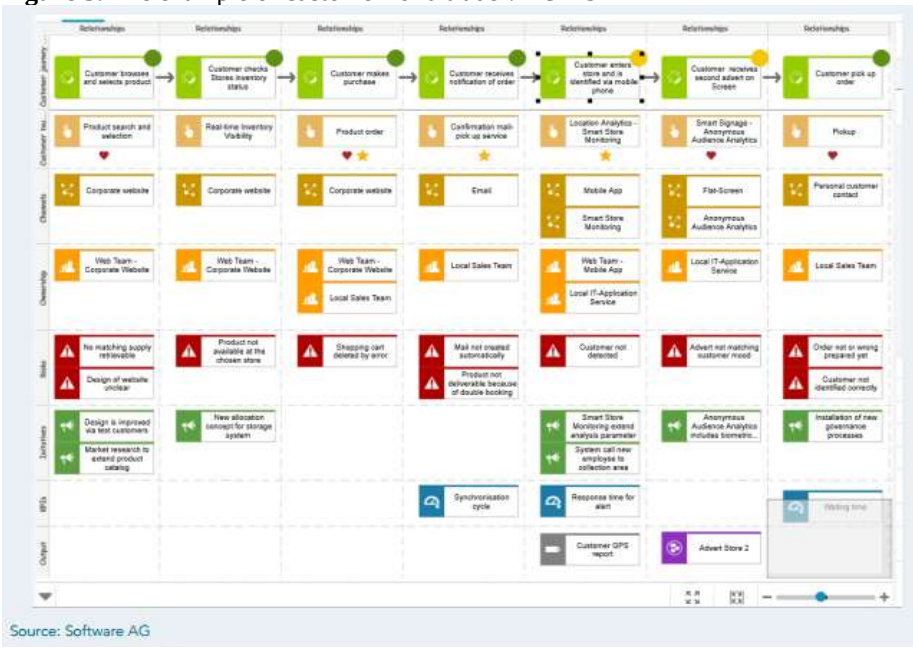


Figure 3. shows a map of journey through the example of a customer for a trader focused on the BOPIS suggestion.

Figure 4. represents the capability card, which shows how the key internal systems should be developed and the capability through the connection of points of the following steps.

Figure 4. Capability card

Figure 5. Respecting the customer

Source: info.softwareag.com

Good governance and the management of business processes help with the way that allows teams to coordinate work and effectively share knowledge through every customer journey.

5. DISCUSSION

The moment when it becomes clear how this critical segment of work is usually and the worst possible – is when something stops working, when there is a problem! The only way that both problems can become benefits and the best indicators of how organized and continuous business process management improves business is – the introduction of a single platform for managing business processes. A repository of business processes set up in this way becomes a virtual company - this is the central place of communication of all segments of the company, all profiles of employees who participate in the management of individual processes in any way, group of processes, business area or the whole company.

6. CONCLUSION

In this paper, we have processed the application of tools in managing business processes, and we defined the problems that appear in the management of business processes between customers, clients, and companies that are in conjunction with tools that monitor the solution and improvement of business process management that is absolutely essential for the functioning of project management in a company that generates profitability and adequate credibility on a yearly basis where we saw that the application of new application software and tools to improve business processes has a major share in problem solving and troubleshooting in managing all business processes in a company that is best reflected through the creation of a final an annual report where, on the basis of the cross-section of the situation, improvement and improvement in the overall business process can be seen, where productivity and efficiency can be seen through the application of adequate software packages that are in operation. This work will definitely serve in the future as one example as well as public and private companies that are in the period of improving the management of their business processes.

7. References

1. Sajfert, Z. and Ćoćkalo, D. (2009-2010). Preduzetništvo, Technical Faculty "Mihajlo Pupin" Zrenjanin University of Novi Sad, 201p.
2. Nikolić M., (2007/2008). Strategijski menadžment, Technical Faculty "Mihajlo Pupin" Zrenjanin University of Novi Sad, 93p.
3. Đorđević D., Ćoćkalo D., (2007/2008). Upravljanje kvalitetom, Technical Faculty "Mihajlo Pupin" Zrenjanin University of Novi Sad, 1p.
4. Bailey J. P.: Internet Economics, Cambridge, 1998
5. Bosilj V.: Upravljanje poslovnim procesima, Zagreb, 2004.
6. Dickens P.: Global Shift Transforming the World Economy, New York 1998.
7. Hill, W.L.C.: Global Business. Boston, 2001.

8. Qualitas d.o.o. Zagreb http://www.qualitas.hr/poslovno-savjetovanje/upravljanje-poslovnim-procesima.html.
9. http://www.evision.hr/hr/Novosti/Stranice/zasto-razumjeti-poslovne-procese-upravljati-procesima.aspx
10. https://pcpress.rs/uvod-u-upravljanje-poslovnim-procesima/

Internet marketing - Start by asking questions

Sanja Stankov[1] Slađana Borić[1] Dragiša Tolmač[1] Aleksandar Rakić[2]
[1] Technical Faculty "Mihajlo Pupin" Zrenjanin University of Novi Sad, Djure Djakovica bb 23000 Zrenjanin, Republic of Serbia
{sanja.stankov}sssanja89@gmail.com

[2] Faculty of applied ecology "Futura", Belgrade, University Singidunum,Pozeska 83 11000 Belgrade, Republic of Serbia
{alexandar.rakic}rakic89@gmail.com

Abstract. Online marketing is moving at the speed of light. To be ready in such an aggressive industry, we must think critically and be tirelessly creative. The content of this paper is intended to make interesting information interesting, memorable and easily shared practically and clearly displayed. People who are engaged in online marketing they encourage us to take action. And anyone who is in any way dealing with marketing, knows how important it is. People love good content. Why? Because it motivates us and inspires us and makes it easier for people to connect with the information that is placed before them. The launch of sterile information in the blank market does not simply pass. Active information is very important that has its effect of periodic repetition at certain intervals where the content is expanded and shared, and the information comes to end users of the content.

Key words: on-line marketing, creativity, action, content,company

1. Introduction

In order to achieve success on the Web, the crucial success factor is the number of visitors that will come to your site. Internet marketing can be confusing, especially if you first encounter it, which is all right. One of the first emotions that a person feels when learning something new is fear, followed by confusion. You will never be alone on the path to achieving a successful goal as an internet marketer. You are not the first to want to learn about internet marketers, nor will you be last, and be sure that the answers you are looking for right now are all on the Internet.

There are many ways to succeed in online marketing, and opportunities to earn money online have never been better. Internet marketing can represent different things for the different people they use on the road to success by combining what they have learned from other, common Internet marketing strategies and their personal formulas. Keep in mind that the Internet has become a big "thief" of time because you can lose a lot of hours by surfing the Internet.

The time spent can be a part of the research or learning process and it is best to limit the time provided for this segment of work in advance. Keep in mind the simple formulas that, through the creation of high-quality and interesting content targeted at target groups, produce the result of the expected effect – purchasing.

2. Be laser focused on your customers

At any moment, organizations have hundreds of options to launch marketing campaigns, from blogging, send newsletters, social networking campaigns, etc. Also, based on the literature studied ([2, p. 62-64]; [3]; [4, p. 4]; [5, p. 4]; [6, p. 167] we believe that we can Zimmerman and Sahlin mentioned [7, p. 11], that social networking sites, social networking sites, blogs, microblogs, forums, collaborative projects (wikis and social bookmark sites) social networks for sharing content, virtual social worlds and virtual online games. marketing mix 4P, which in traditional as well as in digital media consists of a product, price, place and promotion (marketing communication).

Social networking is comprised of people, platform, participation and promotion [8]. Marketing communication is a key factor in the spreading of information and the development of new fields for online income and favorable financial conditions that are caused by the flow of online funds. and promo their product's.Marketing is a strategic indicator to succeed, you need highly focused goals. It's not important whether you're a beginner or an advanced marketer - this fact will always hold true.

If you run without a direction, you'll end up wasting two of your most valuable assets: time and money.The first step is understanding your company's exact needs and goals. Average marketing experts think about campaigns. They work all week, push out a campaign, then start again from scratch next week. To get to the next level, you need to start thinking in systems and build a marketing machine.Marketing is about human-to-human relationships and can happen through any online or offline medium.

3. Companies in the role of providing services and finding clientss

An example of a marketing strategy is certainly a company that was considered to be they must have run robust marketing campaigns, The Eagle Rock Brewery. The Eagle Rock Brewery in Los Angeles has never spent money on print or online advertising. Instead, the company creates a high-quality product that people enjoy. The truth is, Eagle Rock's secret marketing weapon was and has always been a word of mouth that is transmitted from a human being that's what And that's just it. It's Eagle Rock Brewery's reputation that fuels the company's online marketing strategy has gone a step further and is starting to evolve.

What the Internet did was amplify the natural word-of-mouth effect and make conversations stronger. Eagle Rock Brewery has an impressive social media following. Their strategy is simple to keep communication as open as possible with their

communities of customers and prospects. It is a single, focused goal. So, it does not matter if you're running online ads, e-mail marketing campaigns, or retargeting customers on Facebook. What matters is that your company's marketing team sets up the right systems to build genuine human-to-human bonds with the buyers you're trying to reach.

The key to successful Internet marketing is your ability to choose the best possible channels for achieving that end goal.Marketing is very much like public speaking or performing. Even if you're talking to a big group, you need to figure out the best way to connect with your viewers on a individual level. That means understanding who is in your audience as well as what brought them there.

4. Develop your customer personas

There is no such thing as a one size fits all customer. Your buyers are highly diverse and will demonstrate their own unique preferences, personality traits, and needs. Your online marketing initiatives need to reach each and every one of these customer segments. These are the behavioral, demographic, and psychological characteristics of your buyers. Start with the following questions:

1. What are your typical customers' roles? If you're selling a B2B product or service, what are their job titles?

Examples for consumer brands include husband, wife, father, mother, aunt, uncle, grandmother.

Examples for B2B brands include IT manager, marketing manager CEO, business development manager, vice president, etc.

2. Do these individuals have decision-making authority? If yes, do they still consult others? If not, who has the final say?
3. Demographic characteristics (job title, age, education, industry.
4. Where they're located (city, suburb, rural).
5. Key professional attributes: (responsibilities associated with this buyer persona's job, highest job priorities/responsibilities in this buyer persona's direct area of influence, the top problems/pain points they're facing that your company can help solve, several perceived barriers to the above problems, what actions the buyer may have already taken to solve their key problems, who do they report to, years in this role).

Marketing is very much like public speaking or performing, but, whatever you do, don't copy other companies' diagrams.

5. Implementing a facebook campaign in the function of making a big budget

In any campaign implemented by Facebook, most of the brands presented demonstrate a lack through the predicament and encounters misunderstandings and misunderstandings of users who, on the basis of set criteria, are judged and presented to the general public individually. For example, Prigat is an Israel juice company, which injects the element as you would say, the smile of gold is worth it and the money attracts and saves through the realized money placement of funds.

The mechanism was as follows: the fans activate a juicer, which pushes the juice from the orange; Juice is given for charity purposes. The fans were posting their funniest images, which were processed with face recognition technology. The result: over 20,000 people became their own photos, and more than 40,000 oranges were leached into the juice.

An example of a successful and striking marketing campaign is the Opel campaign AdamYourSelf. In December 2015, they invited people to express their personality and creativity in the #ADAMYOURSELF competition. Over the next three months, thousands of people have designed their ADAM online. Through this campaign, Opel has enabled everyone to design Adama (model car) in its own taste and desires. By choosing different shapes and colors of wheels, exterior, mirror covers, roof, foil design, everyone could make customized Adam by simply clicking on different options.

After giving us a chance to try out the role of designers and dreaming of the "ideal" car to our advantage, the Opel team allowed us to capture the car that we arranged by inviting friends, acquaintances, family, colleagues, random people to vote for us and our model Adam, as the most beautiful. This naturally resulted in an organic number of shareholders, invitees, customized Adams, so that many campaigns can appeal to viralism.

Complex applications cost, of course, but not every successful Facebook campaign has required a lot of resources. The most important is in the whole story of the idea - something new, something that will stimulate and retain the interest of visitors and encourage their share. Every single marketing campaign should have a causal relationship to revenue. With so many options, strategies and tactics to consider, how do we really know what's best for our company?

6. Key takeaways

Marketing starts with our customers. Before launching any campaign, we need to understand what our customers want and need. They are always the top priority. Why else does business exist?

When the strategies are timed, they must think beyond the basic metrics, the time of the day, the day of the week, or the month, rather than the perspectives and psychology of the customers. Go from the perspective of marketing analysts to see you as an

ordinary consumer. How do our favorite brands come to communicate with us, what marketing channels do they use?

Each marketing plan requires clearly defined goals.

For example, you can set a goal to acquire new users. Why? Because your business needs paying customers to stay sustainable and grow.

Another goal you can set is to boost your shares and follow on Facebook. Why? Because social media engagement generates exposure for your brand.

7. Marketing strategies in response to key issues in Internet marketing research

A carefully planned strategy is needed. Every marketing plan needs to start with the results that you want to achieve. Think about it:

You would not build a house without a blueprint.

You would not start a business without a business plan.

You would not take a road trip without your navigation system.

The question that always arises when marketing is mentioned is money? How much should I invest? Here's the thing. Online marketing is profit-driven. It's not an investment, and it's not an expense. When executed correctly, it works. We know it works because it's possible to track anything and everything, all the way down to the source of the click that generated the sale. With the right tools and a little bit of creativity, you can prove the value of your marketing. Web analytics is a great solution.

The success of marketing comes only when it comes to right customers at the right moment while they make a decision to buy. During the holiday season, we notice ads for tourist arrangements in our Facebook feed. Coincidence? Absolutely not. Strategic and careful planning? And more than we can imagine. For a random observer, these promotional efforts seem to be coincidental, while with the other side of the computer screen is a marketer who carefully analyzes and responds to your behavior patterns.

It's key to remembering - that timing is absolutely everything. If you are not sure what the company's goals are, in order not to throw darts in the dark, deal with asking questions that require answers before it dives into any marketing strategy: What are the strategic goals of the company? How much can you invest in marketing? What realistic results do we hope we can achieve? A good way to get the answer is to look at brands similar to ours and with similar financial structure.

Before joining the marketing campaign, answer the question your campaign wants to achieve before you are exposed to marketing costs. Even if you just want to try new methods and something new, keep in mind that the timing in marketing is everything. The right message, to the right group of people at the right time, and do not underestimate the power of this statement.

8. Your brand's story

Human-to-human connections are the heart and soul of business. How to tell a story that will tell customers to connect to your brand. If you create it in a quality way, the content can very easily become the best seller you ever had. Everything revolves around a true understanding of your customers' needs and direct communication with their instincts and what drives them.Tell a story that will empathize with customers and make them sit down and think about what they just read on your site or social network. But what kinds of stories do we need to use for business purposes?

What stories will make customers fall in love with your brand and the products you offer them? All kinds of stories, actually.Sponsor who is the hero of your story. Every story is an indispensable individual who is at the center of everything. The hero can be a real person, a fictional character, or even an object. It does not really matter, as long as he or she is someone who changes in the course of the story and goes through some kind of transformation. Although this principle uses most brands in its content marketing strategy, companies often repeat one and the same mistake - their brand performances as the hero of the whole story. Instead of putting special emphasis on customers, how are they who need to get the final value from the content they read.

An example of good practice is Under Armor, whose founder and CEO, Kevin A. Plank, created a product that really solves the problem that athletes often encounter. And naturally, he was very interested in this ambitious venture. Sportswear that remains dry no matter how sweaty? Dream of every professional athlete on planet Earth. Naturally, Plank set up this innovation as a core of his company, stories and overall marketing, bringing Under Armor into a very serious brand within a short period of time.

They know that they really pack the story to their target audience in a really good way. The content they offer is a clear way to overcome the obstacles, shortcomings and, in fact, everything that can be found between their consumers and the goals they set themselves.Give the audience an opportunity to identify with you. People often opt for a particular product or service if they are identified in a particular story, i.e. If they see that someone similar to them has solved some of their problems using those products or services.

Before buying, they first want to really feel that this is something that can surely benefit them. A well-told story can persuade people to decide exactly for your company in the sea of other similar ones, because customers want to see how they can improve their lives and make them feel good. It should be remembered that witnesses sell products.At the end of the day, you're dealing with people - your company is solving problems, alleviating pain points, and providing delightful customer experiences. Revenue is something that happens as a product of a sound business model and a positive customer experience.

Good stories give big voices to small ventures.

9. Conclusion

Believe in what you are doing. Put yourself in the customer's focus by providing meaningful and targeted content. Let the message be simple. Transfer this message to them as a trusted friend. Do not be boring as a warning at the end of the commercials for medicines, but be fun. And finally, have fun while creating that content. You are an artist with a blank screen in front of you. Leave the imagination at will! The goal is to continue attracting customers, but not in an aggressive way of excelling mass marketing, but using meaningful, quality content for targeted customers. If that part of the job is well done, the buyer will come to you, and if you have created something that solves his problems, he will be returned. If you have managed to get a good voice on your company from mouth to mouth, this means that the buyer will bring other potential customers with you.

10. Referances

1. Miodrag Kostić, Internet marketing i Online PR, 2003 - VEZA d.o.o., Beograd.
2. Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! • e challenges and opportunities of Social Media. Business Horizons , 53-68
3. Constantinides, E., & Fountain, S. J. (2008). Web 2.0: Conceptual foundations and marketing issues. Journal of Direct, Data and Digital Marketing Practice , 231- 244
4. Gunelius, S. (2011). 30 - minute social media marketing. • e McGraw-Hill Companies, Inc.
5. https://www.quicksprout.com/the-beginners-guide-to-online-marketing-chapter-1/
6. https://www.entrepreneur.com/article/243764
7. https://knowledge.hubspot.com/contacts-user-guide-v2/how-to-create-personas
8. https://www.socialmediaexaminer.com/facebook-ad-campaign-changes/
9. https://www.sitra.fi/en/articles/10-key-takeaways-wcef2017/
10. https://www.business.qld.gov.au/running-business/marketing
11. https://lorirtaylor.com/72-questions-to-help-you-dig-deep-in-telling-your-brands-story/

Authoring and Publishing Text Documents by means of Linked Open Data Technologies

Evgeny Cherkashin[1,2,3], Alexey Shigarov[1,3], Irina Orlova[2], and Ivan Mikhailov[4]

[1]V.M. Matrosov's institute of system dynamics and control theory of SB RAS,
Lermontov str. 134, Irkutsk, 664033, Russia
{eugeneai,shigarov}@icc.ru
[2]National research Irkutsk state technical university, Lermontov str. 83,
Irkutsk, 664033, Russia
[3]Irkutsk scientific center of SB RAS, Lermontov str. 134,
Irkutsk, 664033, Russia
[4]Limnological Institute of SB RAS, Ulan-Bator str. 3, Irkutsk, 664033, Russia

Abstract. The problem of automating the creation of text documents with the active use of declarative tools and technologies of Linked Open Data is discussed. Software tools for composing HTML5 documents from various sources and presenting the documents in a specified layout, such as a template or a form, are being developed. The sources are web pages and the results of text processing on the server or a client (web browser). Several examples of the application of the developed software are considered.

Keywords: Linked Open Data, automated document composition, document semantic document markup.

1. Introduction

Linked Open Data (LOD) [1] technology has been suggested by W3C consortium to represent the semantic information in the published web content in a way that provides not only the possibility of its processing with software agents (Semantic Web), but also to link all available information into a single semantic graph using relations and global universal identifiers (URIs) of resources. The descriptive capabilities of semantic web technologies, HTML5 document publishing tools, and LOD technologies form an infrastructural basis of authoring and publishing documents. The document is constructed out of individual parts (text and images) that are loaded from other servers with links to relevant resources. The resources represent both the static content and text content as a result of a data conversion algorithm execution. The LOD provide a logical markup for the information presented in a document, informative basis for the different variants of visual representation and interpretation, logical connections with other documents, export information into other documents, procedural processing, etc. An important advantage of LOD usage in information environments is a weakening of the

requirements to the information warehouses: the document itself is a formalized data warehouse. In some extent, this allows reallocation of the time spent on designing the database structure for storage of partially formalized documents to the process of solving a domain problem: the user (developer) markups the document text data with semantic meaning.

The aim of the study is to design an approach to the development of software performing the functions of authoring and typesetting documents marked up according to LOD principles. To achieve the goal, a set of server functions and client (web-browser) JavaScript routines has been developed that allows user to create documents with web browser. The functions support uploading of text fragments from other documents and their generation with server and client scripts. The resulting document is stored in the data warehouse, printed or downloaded to the user workstation as an HTML file. JavaScript implements various document views, e.g., as an input form; text part conversions, e.g., leaning nouns and substituting pronouns.

The created editing tools are the basis of developed digital archives of documents intended for solving problems of creating documentation of university courses, representing research results of DNA/RNA sequencing. The use of LOD data formats, HTML processing tools of web browser and developed technologies allow one to solve a wide class of problems of formation of documentation, ranging from the formation of a meaningful part of the texts through the presentation of stylistic characteristics of the texts and the integration of logical markup into a global data access services.

The development of the tools for document representation on the basis of LOD are aimed at the formation of the prototype of a global environment that supports automation of data transfer between documents of several companies, and creates, ideally, a distributed document processing environment with the principles of social networks. The task is relevant especially in some small business environments with no well-established information flows and relationships between documents. This is mainly due to the dynamism and diversity of the small business. Use of large-scale workflow automation software in the environments of such kind requires large financial expenses for the purchase and continuous improvement of the software product. Moreover, in companies with a deployed a document flow automation, there is almost always a wide class of documents, which are not integrated in the main document flow (e.g., memos, statements), documents with difficult-to-formalize information (e.g., job descriptions) and the documents received by the organizations from the outside having no logical markup (e.g., industry standards).

This work continues the research outlined in [2] in implementing client services operating on the platform of the web browser.

2. Technology of document authoring based on the LOD

The developed software for typesetting of text documents are based on technologies, which implement the HTML5 standard available in modern web browsers. The document is a web page that is composed out of different parts by means of the JavaScript routines functioning within the web browser. To make changes in the content of the page, built-in browser editing tools of web pages (`contentEditable` attribute) are used. The tools are managed by the special libraries (Medium Editor,

https://yabwe.github.io/medium-editor/), which extend the basic functions of the built-in editor with document structure conversion and basic LOD markup.

The text of the document is being marked up with RDFa attributes. The semantic markup defines the logical structure of the document, the domain relations between the entities presented in the document. The LOD markup describes, e.g., the distinct parties in an agreement, subject of the agreement, various formalized data (passport data), the "part-whole" relationships sectioning the text document. In addition to the semantic markup, command structures are embedded in the text of a document. The structures are used for activation of algorithms of a content conversion, for example, some structures activate the inclusion of parts of other document in the current one, and other instructions organize grammatical transformations of sentences. The command structures are recognized more easily than the structures of semantic markup. Their interpretation mainly aimed at algorithmic text processing.

As the basis for the semantic markup of the documents, the results of the open project Dokieli [3] is used. The project focuses on implementation of HTML5 text-editing tools for authoring a scientific publication by several co-authors. Dokieli allows one to markup text with RDFa tags, copying the document in its own repository and make annotations of the text with readers' and authors' comments.

In the markup description of a document, standardized ontologies are used.

Open Annotation (`oa`). The ontology standard is accepted in 2017, its main purpose is to represent the content (an annotation) describing other content. The browser bookmarks are examples of such content. An `oa`-resource (a bookmark) is in two major relations to other resources: `oa:hasTarget`, `oa:hasBody`. The first relation refers to the annotated content with a corresponding resource URIs, and the second one, which can be a multiple relation, refers to an annotation resource, for example, a text description of the annotated resource. The edited document is an annotation to the target document, i.e., at the beginning the documents annotate themselves, and after the completion of the final version of the document, the annotation describes the generated immutable copy, e.g., the PDF file of the document.

Friend-of-a-friend (`foaf`) describes information about agents: physical and legal entities, as well as software agents. This ontology is widely used to represent relations between agents in social networks.

Provenence (`prov`) provides a vocabulary of terms and relations describing the origin of the information. This ontology is a convenient way to refer the document's source parts originated from other documents. The `prov` ontology is a basis of informational flows description in documents and their mutual relationships.

Dublin Core (`dc`) is used to describe the elements of authored annotations, for example, section headings, document title, document types, various descriptions contained in the document. In contrast to `oa` this ontology describes the documents mainly at the metainformation level.

DBPedia resource (`dbr`) is the namespace of the objects (resources) of Wikipedia. Ontology `dbr` is a convenient way to refer to specific instances of objects and concrete classes, for example, the passport of a person (`dbr:Passport`) or a city.

Schema.org (`schema`) represents the objects that are recognized by the site scanning agents of Google, Yandex, Yahoo, etc. This ontology is used in the case if none of the above specialized ontological have a relevant term or a relationship denotation. This ontology contains useful abstract relationships like "part-whole" (`schema:hasPart`),

"result of a creative activity" (`schema:CreativeWork`). The relations are used to describe the sectioning of the document, i.e., its hierarchical structure.

In addition to the above-mentioned ontologies, the standard ontologies (RDF, RDFs, RDFa, XSD) and ontologies from the NEPOMUK project (**`https://userbase.kde.org/Nepomuk`**) describing objects stored in the index of full-text digital archives are used as well. They allow one to describe detailed meta-information about the creative works, e.g., files, images, documents, document sections. In our research, there was a labor-intensive stage devoted to the search ontologies to represent documents by subject areas. The resource Linked Open Vocabularies (**`http://lov.okfn.org/dataset/lov/`**) happened to be a very useful reference book as well as support of the community of researchers, who develop this resource.

The architecture of the system is presented in Fig. 1 (all relations shown in the figure are bidirectional). The document is downloaded from a server where it is composed from a content stored in a database or a file system, results of an algorithmic data conversion of other sources, and elements of a template that includes the user view interface elements and interpretation modules on the displayed page. In the formation of the final form of the document, interpretation modules request the required data and resources on the source server or other Internet servers.

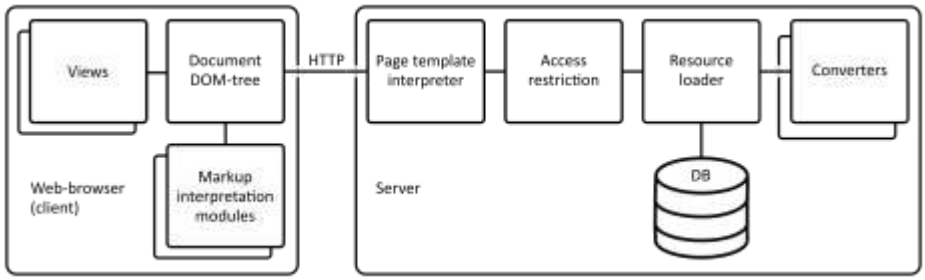


Fig. 1. The general architecture of the document publishing system

Interpretation modules of the semantic markup are implemented with client-side JavaScript. They are executed as soon as the main content is loaded by the browser. The modules scan the tree structure of the document, recognizing the conditions of their activation. If the scan was successful, the body of the module is executed, changing of the document content (tree structure). The document composition finalizes as soon as all conditions were met and all commands activated.

A database of the system is a set of tools for storing and serving multi-format data. The body of the document loaded at the first stage is stored in XML (XHTML) files in the file system on the server. The file system is wrapped with a document versions tracking layer to monitor the document dynamics and create integrated backup copies. This level is quite easy to be implemented using modern version control systems like GIT or Subversion.

In order to organize standard SPARQL data source access and full-text search, storage components for the logical layer and related data as a graph of triples have been developing. ClioPatria [4] and Jena are the component implementations. The first

system is interesting because it is implemented entirely with the programming languages Prolog and C. It supports several formats of compact storage of data, and tight integration of triple data with the runtime and language environment of Prolog. A regulated access to the triples from JavaScript is supported using special Pengines protocol [5]. There are a number of implementation libraries for the protocols for popular programming languages, including client-side JavaScript, Java, Python. As an alternative, Java-based Jena library can be used for the service realization.

A full-text index is provided by Elasticsearch indexing engine. The implementation of the index service is simple enough because any RDF graph is representable as JSON (JSON-LD) document, JSON is the main format of storage of indexed information in ElasticSearch. For the representation of the documents to the user as a search result on the page, some triples must be marked as representative of the result. Elasticsearch has the means of fuzzy comparison of terms that allows construction of relevant information search engines. The general technical aspects of the module functioning are discussed later in the application section.

3. Analogs of the software system under development

One of the projects related to semantic markup control is Semantic MediaWiki; it extends a Wiki engine used for the site content representation. The text of the document is marked up with RDFa attributes with wiki tags. Semantic annotations allow development of search engines for Wiki pages that takes into account the semantics of the marked text [6].

Unlike Semantic MediaWiki, in OntoWiki project, similar results were obtained implementing different idea. OntoWiki is based on the primary usage of the logical description of information in a semantic network. This logical structure is edited by means of the software generated user forms for the known terms in a vocabulary. The user can change only one text property of LOD `lod:content`, which, in General, contains HTML texts. The HTML markup is not tied to the logical structure of the displayed object (the subject). The `lod:content` text is edited with OntoWiki built-in WYSIWYG editor. The OntoWiki project is aimed at support of social network technologies based on Linked Data [6].

As was mentioned above, used format of the markup of the documents is a further development of the results of the Dokieli project located at web site **`https://github.com/linkeddata/dokieli`** [2]. The project is a WYSIWYG-editor of LOD marked HTML pages, display style sheet (CSS) and built-in subgraph (its content is not visible on the page) presented in different formats (TTL, N3, JSON-LD, TriG). The user is able to add new RDFa markup with editor tools, in this case, the function of our interest is implemented in a general form. In addition, user authentication functions are implemented according to the WebID standard with client-side JavaScript. The system extensively supports commenting (text markup) with Open Annotation (`oa`).

The annotation engine is integrated with `gitter.im` service, which shows readers' comments in the form of a dialogue. All the information is stored either on servers that support the protocol Solid (**`https://solid.mit.edu/`**), or locally in the database of

a browser. The edited document can be easily moved (copied) from one Solid store to another. The resulting texts can be downloaded to the user workstation as XTHML documents. Thus, Dokieli is a system supporting distributed text content editing, which fully implements the interaction with other LOD resources, and all these functions are implemented exclusively by means of the client-side browser JavaScript.

There is also a huge class of browser document editors focused on the creation of scientific publications, see, e.g., **http://substance.io/**. In our study, the development of tools is aimed at automating the markup of a document based on the analysis of document changes proposed in [2], and specifically in this work, at implementing one of the infrastructural problems connected with a creation of the documents on the platform web browser controlled with LOD.

4. Applications of the technologies

The developed tools for typesetting documents are used in applications.

4.1. The texts of the curricula of the universities

The Ministry of Education and Science of the Russian Federation after a series of experiments started the large-scale deployment of the Bologna process in the educational environment of the Russian Federation. One of the goals to be solved in the introduction is the transition to competence-oriented representation of the requirements to the pedagogical process results. The reform affects all aspects of the process, including the system of classification of professions, introduction of "programs" and "directions", specialization by skill level (bachelor, master, etc.), introduction of applied bachelor degree, list of courses, goals and objectives of the courses agreed with the competencies specified in the Federal State Educational Standard, forms of conducting classes, e.g., the introduction of interactive studying forms, the distribution of lectures and practical sessions, etc. The existing documentation of the courses is supplemented with new forms of mandatory documents: annotation of courses and the basis of assessment means (BAM), which is a set of examination questions, tests and test problems used for assessment of students' competences. In addition, the university managers, with the aim of improving the quality of educational services, introduced its own additions to the form, content and requirements to the documentation typesetting, in particular, to the quality of the conversion to HTML for publication on the website of the university.

For the minimum fulfillment of the requirements of the university management for each course the instructor is required to design at least three documents: working program of the course, annotation and BAM. Thus, this set of documents is prepared theoretically for each following possible combination: the university, department, specialty, direction (profile), program, qualification level (bachelor, specialist, undergraduate, graduate), academic or applied version of the qualification, form of study (full-time, part-time, evening, part-time, etc.). Each combination is presented in the curriculum of the university from which data is annually or twice a year should be

synchronized in the work programs. The last five years showed the level of forethought of decisions taken by the managers of the ministry in terms of requirements to the presentation of a course – four generations of the federal education standards (FES-1,2,3 and 3+) have been developed – for each generation the documents in the appropriate "new" form should be prepared. The task of developing new handbook now seems to be simpler than before on the background of the document set preparation complexity. The raw statistic shows that most instructors are not able to cope with high-quality paperwork in the provided time constraints, resulting in the necessity to departments to hire secretaries, whose function is to bring documentation to the required quality level.

The solution to this problem is supposed to be in developing a software system that allows collecting the texts of the work programs, annotations and BAMs from separate parts: a list of competencies and curriculum; the subject and module contents, whose text parts shared between different versions of documents. The title pages are generated from data of curriculum and designed templates. Microsoft Word and Excel are commonly used for this task with a built-in VBA, but in this case it is clearly not enough.

Consider the scheme of representation of a marked up work program in our system.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
        "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html lang="en" xmlns="http://www.w3.org/1999/xhtml"
      xmlns:taa="http://irnok.net/engine/rdfa-manipulation"
      xml:lang="en" metal:define-macro="page">
    <head> <!-- Connecting stylesheets and modules -->
    </head>
    <body prefix="rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns# ...
                           foaf:http://xmlns.com/foaf/0.1/ imei:
imei.html#
                           course:
https://irnok.net/college/plan/01.03.02-16-1234-2461_1%D0%BA_PB-
SM.plm.xml.xlsx-%D0%911.%D0%92.%D0%94%D0%92.3.1.html#"
         resource="#post" typeof="schema:CreativeWork sioc:Post
prov:Entity">
        <!-- The application control panel -->
        <main lang="en" resource="#annotation" typeof="oa:Annotation"
            id="main-document-container">
            <div property="oa:hasTarget"
                   resource="#course-work-program"></div>
            <article property="oa:hasBody"
               typeof="schema:Article foaf:Document
curr:WorkingProgram"
               resource="#course-work-program" id="main-document">
              <div taa:content="imei:title-page"></div> <!--The title
page..
              <div taa:content="imei:neg-umc"></div> <!--The approval
page..
                <section id="contents" class="break-after">
                    <h2 class="nocount c">Table of Contents</h2>
                    <div id="tableOfContents"></div>
                </section>
                <section id="course-description"
resource="#description"
                 property="schema:hasPart"
typeof="schema:CreativeWork">
                 <div property="schema:hasPart" resource="#purpose"
                          typeof="dc:Text cnt:ContentAsText" >
                        <div property="cnt:chars"
```

```
datatype="xsd:string">
                              <h2 property="dc:title"
datatype="xsd:string">The goals and the objectives of the discipline
(module)</h2>
                              <p>
                                  The purpose of the discipline
"Programming Technologies" studying is the development of ...</p>
                          </div>
                      </div>
                      . . . . . .
                      <div property="schema:hasPart" typeof="dc:Text
                             cnt:ContentAsText" resource="#volume">
                          <div property="cnt:chars"
datatype="xsd:string">
                              <h2 property="dc:title"
datatype="xsd:string"> The amount of discipline (module) content and
training activities (divided by form of training)</h2>
                              <div taa:content="course:time-
distrib"></div>
                          </div>
                      </div>
                      . . . . . .
```

In the presented example, the key structures are highlighted with bold font. Let us comment these structures.

1. The page displayed to the user is the annotation of the document "`#annotation`", with both the annotated content and the annotation text being the same at the stage of formation of the document - resource "`#course-work-program`". In LOD all the resources are global. In this case, this is achieved by substitution of the default namespace - the full URI of the current page - on the left hand side of the name of the resource.
2. The title page and the special approval page are inserted from the page templates prepared in "`imei.html`" (information about the Institute of Mathematics, Economics and Informatics of Irkutsk state University). The course data and the name of the profession are inserted into the template from the context of the document. All the key static templates of the courses can be placed on one template resource page.
3. The text is divided into sections, and wrapped by tags `<div>` and `<span>` with the relevant RDFa structures. Analysis of the experience of the developers of LOD resources has shown that for the formation of the relationship is sufficient to use RDFa tags `property`, `typeof` and `datatype`. The usage of `rel` and `about` should be avoided. This makes the structure of semantic markup to be stricter by reducing the number of entities.
4. The command `taa:content` adds text from another page to the current document. The address of the included text is formed by the interpretation of the parameter of the command, for example, `taa:content="course:time-distrib"` adds the text of the table of distribution of time spent between the types of classes (lectures, practice, laboratory works, etc.). To generate such tables, a server of web pages of these courses has been developed. Each page represents a course/module presented in a curriculum. The page content is defined by the structure of its URL and rendered at the time of the first loading to the browser. The text templates are identified with `id` attribute.

In edit mode of the work program (document), all tags having the combination of attributes `property="cnt:chars"` and `datatype="xsd:string"` are converted to editable text. For clarity the document is displayed to the user without `taa:-inclusions`, and these inclusions are indicated with special tags and styles. At the finishing time moment, the edited text stored on the server file system. After that, the user can commit the changes and synchronize the text of the saved document with a repository.

Project source code is available in modules of the Python programming language at the following address: **`https://github.com/isu-enterprise`. The main module is `isu.college`**. A partially functioning example of the document rendering (in Russian), which could be edited in the browser, is located at **`https://irnok.net/college/doc/ProgrammingTechnologies.html`**.

4.2. Designing legal documents

Students of the Institute of Mathematics, Economics and Informatics of Irkutsk State University developed an application for creating legal documents, non-standard form of judgments of judicial officers and notarial documents. The markup documents in this project are similar to the one used in the previous example, a number of ontologies for representation of significant terms in the legal domain were added: `fibo` (The Financial Industry Business Ontology), and formal specifications of identity documents `acrt` (A Certification Ontology).

`Fibo` ontologies allow specifying the role of entities in the document, e.g., for a general power of attorney, who is the principal (truster) and who is trusted (the text is in Russian):

```
<p>Я, <span
       property="fibol:designatesSignatory bibo:owner"
       typeof="fibol:Signatory foaf:Person dbr:Principal"
       resource="#principal"><span property="foaf:name" id="signatory-
name" datatype="xsd:string">Иванова Елена Викторовна</span>,<span
property="adoc:hasPassport" resource="#signatory-passport"
typeof="acrt:Certification"> <span property="acrt:qualification"
resource="dbr:Passport">паспорт</span>…
```

This example expresses the fact that the principal (`fibol:designatesSignatory`) is being certified (`adoc:hasPassport`) with a passport (`acrt:qualification dbr:Passport`).

The application has a mode for the document-editing in a form view to be filled in. The JavaScript module searches for the tags containing a set of attributes `datatype="xsd:string"`, `class="edit"` and converts the tag text into an input field, while the main text remains static. The values of the form input fields stored as tag texts or as values of tag `content` attribute, which is allowed by the RDFa standard. The attribute allows, e.g., storing peoples' names in the nominative case, and the displayed text to be in a suitable declination.

Harmonization of grammar is implemented using the attributes `id`, `data-` and `class`. The value of the `class` attributes equal to `"disp"` indicates that an edited string identified by `id` attribute is substituted as the text in the tag. The attributes `data-`

m (masculine) and data-f (feminine) specify a variant of a word to be substituted according to the genus of the related noun in the text. The attribute data-case specifies the declination (the case) of the words of the text to be substituted. The grammar transformation algorithms are implemented on the server. Here is an example of the previous text view of the power of attorney with grammatical structures.

```
...проживающ<span id="signatory-name" data-m="ий" data-
f="ая">___</span> по адресу:...Подпись <span id="signatory-name" data-
case="gent">___</span> удостоверяю...
```

Copy functions and document data substitution from one document to another has been implemented. Further improvement of the software is a construction of a workflow graph based on analysis of data transfer between documents. The address of the application source code development page is **https://github.com/isu-enterprise/isu.aquarium**. A working example will be presented soon at **https://irnok.net/aquarium/attorneyletter/0001.html**.

5. Further development of the technologies

At present, a large amount of information is published in tabular form in unstructured sources, e.g., government statistics, financial and business reports, sets of open scientific data, social media. The popular data representation formats of the sources are PDF, raster scans, DJVU, Word and Excel formats. These sources of data are a valuable resource for the document authoring. For example, the majority of regulations and education standards of Ministry of Education and Science of the Russian Federation are published in PDF format in scanned form. To be able to use data from these sources, it is necessary to extract the data and transform it into marked-up text or into tables of a relational database. The authors develop technologies for conversion, analysis, interpretation, cleaning and tracking the provenance of the tabular data from unstructured sources, the conceptualization of their natural-language content to provide a solution for this important task. We created an organization Cells Research Group at Github (**https://github.com/cellsrg**), where two projects have been implemented: a system of detection and recognition of tabular structures in PDF documents (**https://github.com/cellsrg/TabbyPDF**); a rule based semantic structure analysis of Excel spreadsheets (**https://github.com/cellsrg/cells-ssdc**). The recognition result of the tables is the data in a relational table [7].

LOD markup for the tabular data taken from the unstructured sources is described using the ontology qb (Qube ontology), which allows presenting the data as a hypercube. For each table cell, in a general case, a coordinate of the cell in the cube and the properties of the stored values, for example, the data type and units have been specified. The qb markup allows direct algorithmic processing for the table, i.e., without additional analysis of the structure.

The results are expected to be applied for automation of data publication of research results in the field of high-throughput sequencing of DNA/RNA carried out at the Limnological Institute (LIN) of SB RAS. In order to prepare the data and paper publication, the text and tables are to be additionally RDFa marked up with formats supported by biological LOD data servers, e.g., BIO2RDF (**http://bio2rdf.org/**).

The implementation of this service for LIN SB RAS, the researchers will integrate data on the microbiome of Lake Baikal in the global research environment of species of plants, animals and micro-organisms similar to [8].

Another important task of analysis of unstructured information is to automate the layout and the structure recognition of similar documents according to a template. The task is relevant for universities, where a large amount of material in the form of training programs and handbooks are accumulated. The material could be converted into various forms of preparation, such as training programs on the modern standard FES-3+, e-courses of Moodle, etc. In addition, the universities of Russian Federation are developing intelligent systems for training automation where the material to the student is given depending on a predefined general scenario and the results of the assessment of the current knowledge. Marked up methodological and documentary material could be used for partial automation of the process of populating the module databases of e-learning systems.

6. Conclusion

The report discusses the utilization of the Linked Open Data (LOD) technologies for solving problems of creating text documents with the editing and programming means provided by modern web browsers. Thanks to the LOD notation, data integration of documents and web applications into a single resulting text document is being implemented.

The editing and typesetting of the documents is carried on using algorithms that interpret the relation between the elements of document's HTML tree. The relations are described by means of semantic web and LOD, i.e., with RDFa structures and commands defined in the attributes of HTML tags. The layout algorithms are activated in the presence of corresponding conditions in the nodes of the document tree. Each algorithm changes the tree structure, constructing the final content and layout of the document. The edited documents are stored on servers that support HTTP Protocol (GET and POST requests). This technology is very scalable even for shared hosting environment supporting only PHP. Test storages are implemented for different programming languages.

A few examples of the use of the developed tools have been presented in an educational environment, preparation of legal documents. Further development of the project is carried out in several directions: a) improvement of the means of typesetting documents, b) the automation of the document structure recognition based on the analysis of its changes [2], c) the implementation of domain applications, d) the collection and analysis of user activity and requirements, d) development of a regulated access to documents. The development is aimed at the implementation of a software environment for global electronic document workflow automation, allowing individuals and organizations to share document data, logical structure which is difficult to formalize within the framework of traditional approaches.

7. Acknowledgments

The results obtained with the partial support of the following projects:

- Irkutsk scientific center of SB RAS No 4.1.2;
- The Council for grants of the President of Russian Federation, state support of leading scientific schools of the Russian Federation (NSH-8081.2016.9).

The results obtained with the use of the network infrastructure of Telecommunication center of collective use "Integrated information-computational network of Irkutsk scientific-educational complex" (**http://net.icc.ru**). The authors are grateful to the community of Linked Open Vocabularies (**http://lov.okfn.org/dataset/lov/**) resource for assistance in the search for formalizations of subject areas (ontologies).

References

1. Bizer, Ch., Heath, T., Berners-Lee, T. Linked Data – The Story So Far. International Journal on Semantic Web and Information Systems. 5 (3): (2009) 1–22. doi:10.4018/jswis.2009081901. ISSN 1552-6283.
2. Cherkashin E., K.Paskal, I.Bychkov et al An Ontology Polysystem Driven Technique for Semantic Markup for Edited Documents Procs. of International conference on Applied Internet and Information Technologies. University of Novi Sad, Technical Faculty ”Mihajlo Pupin”, Zrenjanin, Zrenjanin, Serbia, October 24, 2014. pp. 201-206.
3. Capadisli, S., Guy, A., Verborgh, R., Lange, C., Auer, S., Berners-Lee, T.: Decentralised Authoring, Annotations and Notifications for a Read-Write Web with dokieli, Procs of ICWE international conference, 5-8 June, 2017, Rome, Italy. (to appear) Preprint URL: **http://csarven.ca/dokieli-rww**
4. Wielemaker, J., Beek, W., Hildebrand, M., Ossenbruggen, J. ClioPatria: A SWI-Prolog infrastructure for the Semantic Web. Semantic Web 7(5):529-541, 2016, DOI: 10.3233/SW-150191
5. Lager, T., Wielemaker, J. Pengines: Web Logic Programming Made Easy. Theory and Practice of Logic Programming 14(4-5), 2014, DOI: 10.1017/S1471068414000192
6. N.Heino, S.Tramp, N.Heino, S.Auer. Managing Web Content using Linked Data Principles – Combining semantic structure with dynamic content syndication. Computer Software and Applications Conference (COMPSAC), 2011 IEEE 35th Annual. pp. 245 - 250. URL: **http://svn.aksw.org/papers/2011/COMPSAC_lod2.eu/public.pdf** (access date: 30.05.2013).
7. Shigarov A. Rule-Based Table Analysis and Interpretation // Communications in Computer and Information Science. Springer. 2015. Vol. 538. pp. 175-186. URL: **http://link.springer.com/chapter/10.1007%2F978-3-319-24770-0_16**.
8. Dalamagas, T., Bikakis, N., Papastefanatos, G., Stavrakas, Y., Hatzigeorgiou A. Publishing Life Science Data as Linked Open Data: the Case Study of miRBase Proceedings of the first International Workshop On Open Data, WOD-2012. URL: **https://arxiv.org/abs/1205.2320**

An Environment for Metagenomic Analysis

Evgeny Cherkashin[1,2,4], Alexey Shigarov[1,2], Fedor Malkov[1,2],
Kristina Pascal[2], and Alexey Morozov[3]

[1] V.M.Matrosov's Institute of System Dynamics and Control Theory of SB RAS,
Lermontov str. 134, Irkutsk, 664033, Russia
{shig,eugeneai}@icc.ru
[2] Irkutsk scientific center of SB RAS, Lermontov str. 134,
Irkutsk, 664033, Russia
[3] Limnological Institute of SB RAS, Ulan-Bator str. 3, Irkutsk, 664033, Russia
[4] National research Irkutsk state technical university, Lermontov str. 83,
Irkutsk, 664033, Russia

Abstract. Metagenomic analysis allows describing microbial community with a previously unavailable precision, but requires considerable computing power for solving bioinformatics problems and participation of domain specialists at the stage of the result interpretation. This complicates the implementation of the analysis in a broad biological practice. The development of a domain user-friendly software environment for storage and analysis of metagenomic data has been started. The usage of a dataflow programming system for representation of metagenomic analysis and the schema for a SQL database for storage of the metadata are considered as units of the environment.

Keywords: bioinformatics, metagenomic analysis, Big Data.

1. Introduction

In the last decade, thanks to the invention of next-generation sequencing (NGS) methods and their introduction in practice of research of biological systems a field of research of molecular genetics, namely metagenomics, has been arisen. Its basic principle is that the object under investigation is not a separate microscopic organism, but their communities (microbiomes). The sampled probes stand out with a total DNA sequencing data over the whole set of genes of all microorganisms in the probe. That is, the studied object is the microbiome as a whole, not only those organisms which can be cultivated in laboratory conditions or identified with microscopic or microbiological methods.

Metagenomics allows us to describe a significant number of new groups on all taxonomic levels, broadening the field of view of the world science. A characteristic example is the recently discovered group CPR (*candidate phyla radiation*). No CPR is isolated in a culture at the moment. According to genomic data, its representatives differ in the set of ribosomal proteins, the absence of certain key metabolic pathways and the presence of self-splicing introns in genes 16S rRNA [1]. Phylogenetic analysis indicates

that this group is a sister to all other bacteria, and the level of divergence is not inferior to bacteria, not to mention the eukaryotes [2].

There are two main types of metagenomic studies. The first one, which is simpler, is called *analysis of the amplicons*. In this case a specific taxonomic marker is amplified and sequenced. The marker is universal for the studied species. Usually, the sequence of the small subunit of ribosomal RNA is used as the marker, as this gene is widely used in phylogenetics. The gene is available in numerous reference sequences. For example, the release 128 of widely used in amplicon analysis SILVA database [3] contains 645 151 unique rRNA. The reads obtained from the DNA sequences extracted from the sample under investigation are compared to the sequences in databases, attributing them to a particular taxon of a taxonomic level, obtaining information about the diversity of the microbiome in the studied environment.

The second approach is known as *metogenomic shotgun method*. It is based on sequencing the whole DNA sample instead of the specific locus. With sufficient coverage, this approach allows describing the taxonomic composition of the community, as well as the genes of functional or structural proteins presented in the representatives of the community, including viral ones [4]. On the basis of metagenomic data, metabolic interactions in individual microbiomes can be determined using the databases ePGDBs (environmental pathway/genome databases) [5]. In several works, full genomes of individual species were isolated from metagenomic dataset reads [6].

In recent years the amplicon analysis was applied in microbiome studies for different environments of lake Baikal. The researchers of the Limnological Institute of SB RAS described the under-ice bacterial communities associated with blooms of diatoms [7] and bacteria in photic layer during spring [8]. Bacteria inhabiting the Baikal sponges were studied as well [9]. Finally, the bacterial communities of bottom sediments in the areas of hydrocarbon yields [10,11] were investigated.

In order to carry out the metagenomic studies the significant computational resources and bioinformatics skills are required for data processing and interpretation. The software used for analysis of amplicons includes various library modules of sequence processing, for example, Mothur [12], USearch [13], statistical packages and development environments of data mining algorithms, e.g., R (**https://www.r-project.org**). In order to carry out the studies of metagenomic data, the specialists are required to be able scripting the command shell of an operating system (Linux, Windows), running programs in a distributed computing environment and cluster computing systems, and programming with general-purpose languages, usually R or Python.

Another important problem is the organization of a centralized data storage and providing the efficient regulated access to the data for the users. At the moment the staff of LIN SB RAS conducted numerous amplicon research of the different ecotopes of lake Baikal, the data were collected for several years. There are no strict rules of the storage policy of input, intermediate data and the obtained results. Comparison and integration of data from different studies is also complicated due to its heterogeneity, resulting from the use of various software. The implementation of a system for storing input data, metadata, and results of metagenomic studies in a unified form will simplify the integration of results from different studies and the comparative analysis.

The goal of this study is a software environment development for supporting the processes of new-generation sequencing with organizational, informational and computational resources.

2. The domain analysis

Domain analysis showed that the problems solved in the bioinformational part of metagenomic analysis, together with NGS itself, are well represented within the paradigm of Big Data. At the moment, the scientific community developed data formats for representation and storage of metagenomic information, algorithms and software modules including distributed and parallel implementations on cluster computing systems providing different stages of data analysis.

The solution of the problems within the Big Data paradigm requires the biologist to have software development skills to be a professional programmer in bioinformatics. In order to carry on the analysis of each probe, biologist is to construct and execute a separate program script or perform stage-by-stage execution manually to control each step's results quality. This approach significantly slows down the process of obtaining the final result.

The proposed organization of studies is based on the creation of an information-computational environment that allows one to design and execute scenarios, giving the input data in various formats from various sources, e.g., files, databases, servers of metagenomic information. The environment must also support a cloud storage for intermediate data and the obtained results. A collaborative project of LIN SB RAS and ISDCT SB RAS is devoted to the construction of the environment for the research support. The following problems are to be solved within the project.

1. The subject area and its functional modeling. The classes of functions (problems) are being recognized and presented in the form of software modules. Modules form scripts of problem solving, network graphs of modules connected by data transmission.
2. Metadata descriptions of the modules and structures of input and output data. At this stage, it is necessary to deal with the problem of integration with external information and computational resources. In this case, the standards and standard means of data modeling like ontologies are of critical usage.
3. Decomposition of the input/output data formats and implementation of subsystems of their transformation, accumulation, storage and effective (according to the criteria of time and computational complexity) regulated access.
4. Construction of virtual executional environments and software interfaces for modules, whose source code is inaccessible due to the lack of the source code or licensing restrictions.
5. Development a customized user interface for high level control of the scenario executions. At this stage, a visual programming with the user interface for script development and execution is required to provide flexibility for managing computational processes by domain specialists.
6. Development of subsystems of visualization and interpretation of obtained results, including the modules for interpretation of the process of metagenomic analysis.

3. Dataflow representation of the domain

A popular approach to the representation of the computational process is dataflow programming [14]. The data flow programs are constructed as a combination of the executable modules. The modules receive input data, process it, and produce output. The approach is being developed since the 1970-ies.

An example of usage of the script construction system under development is shown in Fig. 1. The figure shows an example of an initial stage of a computing process of analysis of the amplicons.

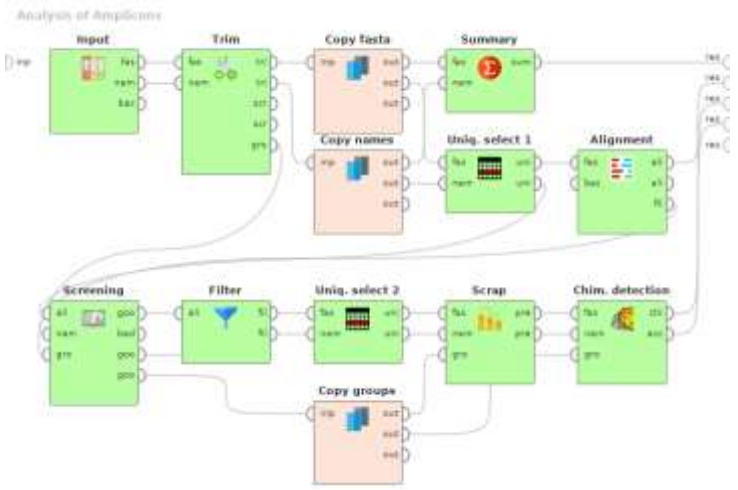


Fig. 1. The initial stage of the metagenomic analysis represented as a data flow.

The presented script was constructed by means of the software package Rapidminer Studio (https://rapidminer.com/) supplemented by our extension module for description of the amplicon analysis stages. The scenario includes the following operations:

- definition of a research project as a set of input files containing the sequencing data in a directory (module “Input”);
- trimming reads (module “Trim”);
- the module “Summary” is used for visual quality analysis of the results of the previous steps;
- the reduction of the volume of input by the removal of insignificant information, for example, overlapping sequences (modules, “Uniq. select …”);
- alignment of sequences to the reference database (module “Alignment”);
- filtering sequences according to specified criteria (module “Screening”);
- removing alignment columns based on specified criteria, for example, empty columns (module “Filter”);
- removal of sequences containing sequencing error (the module “Scrap”);
- detection of chimeras (module “Chim. detection), etc.

The diagram shows service modules of RapidMiner Studio, which are necessary to distribute the same type of information between modules (e.g., “Copy groups”). The

necessity of introduction of such modules is a feature of Rapidminer Studio; it supposes that in a general case the modules make changes in the data under processing without copying it.

Each module receives file names as input and creates new file set as the result. The operation of the module depends on the parameters specified by the user via user interface of each module. The results of the script are sent to output ports and displayed by the Rapidminer Studio visualization subsystem in a convenient form to the user. The system supports presentation of a scenario as a new block with its input and output ports, as well as a cloud storage and execution of scripts, creating distributed computing environment. Rich feature set of Rapidminer Studio and various services provided by its developers were the main reason for choosing this system as a development environment for informational-computation resources of the project.

4. The database supporting metagenomic analyses

Assessment of world experience of organization of scientific research in the field of Data Science showed that the use of cloud technologies is a necessary basis for the interaction organization of the researchers. A specialized data storage should be a unit of the environment to ensure effective user access and computing processes to the data of research.

Database for microbiome-based metagenomic analysis data (Fig. 2.) provides the storage facility on all the stages of the microbiome studies from the probe sampling to the publication of the scientific meaningful results. The scheme in the Fig. 2 represents database structure as an ER-diagram. The scheme contains data about sampling, analysis of physicochemical and biological parameters of the probes, the sequencing results, the applied equipment and software, taxonomic databases, methods of the analysis of the collected material, publications of the obtained results and the participated researchers. It also allows us to store the processing scripts of analysis of metagenomic data, including software tools, commands, and configuration files. The latter function allows one to save the state of the computational process and restart it from the specified point.

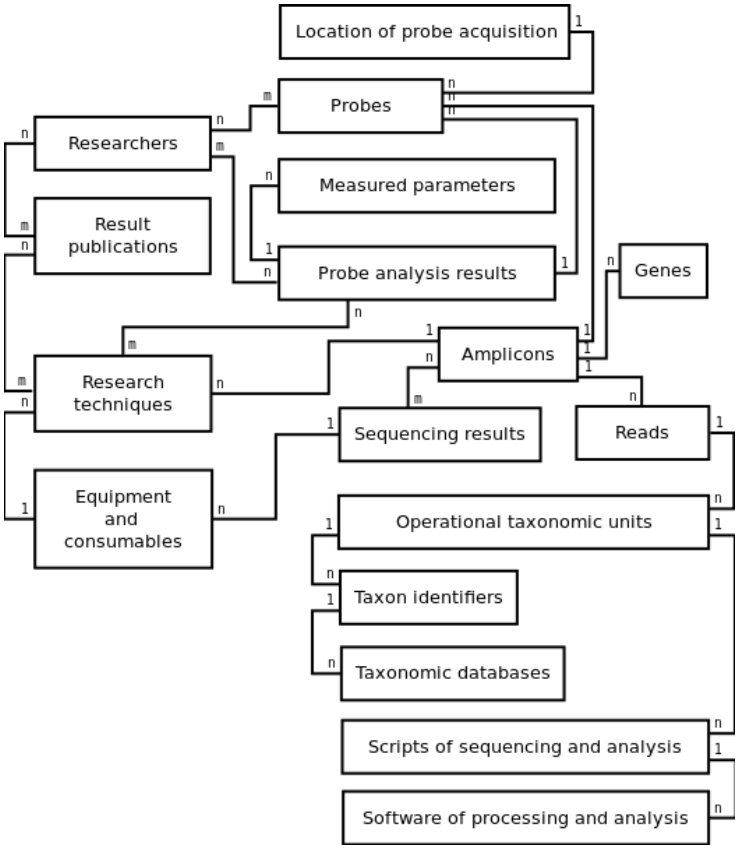


Fig. 2. A general schema representation of the database for microbiome research based on metagenomic analysis

The model is implemented by means of Django framework (**https://www.djangoproject.com**). The framework supports automatic definition of rational table structures representing many-to-many relations, and generation a customizable interface for the administrative panel, allowing testing the developed model. These same tools are used for the implementation of the project web site.

Cloud-based storage and dedicated storage of the metagenomic data will allow one to create online services for joint processing of sequencing data from different studies by specialized software and to publish information in the Internet. In order to achieve this goal, the following must be carried out:

- a software interface implementation for data access;
- filling in the database with information collected and processed as a result of studies of the microbiome of lake Baikal in 2009-2015;
- realization of the scenario design and execution to support the metagenomic data analysis in a distributed computing environment.

5. Conclusion

Modern problems of development of a distributed software environment for the implementation of organizational, informational and computational resources for scientific microbiological studies based on metagenome analysis are presented in the article. A generalized domain model of system-level is conducted, as well as the requirements are stated to the development environment and problems to be solved. A computational model of the process of analysis of the amplicons is being constructed and implemented. Aspect of informational supply of the computational process is represented by realization of the problem of cloud storage for computing processes (scenarios of metagenomic data processing), as well as by construction of a database for storing input and intermediate data, results of the scenarios execution. The database is used as a basis of an information portal construction for processing metagenomic data and presenting the results of scientific community.

6. Acknowledgments

The results obtained with the support of the following projects:

- Irkutsk scientific center of SB RAS No 4.1.2;

The results obtained with the use of the network infrastructure of Telecommunication center of collective use "Integrated information-computational network of Irkutsk scientific-educational complex" (**http://net.icc.ru**).

References

1. Brown, C. T., Hug, L. A., Thomas, C. B. *et al*. Unusual biology across the group comprising more than 15% of domain Bacteria. Nature. — 2015. — Vol. 523. — P. 208-211.1.
2. Hug, L. A., Brett, J. B, Anantharaman K. *et al*. A new view of the Tree of Life. Nature Microbiology. — 2016. — Vol. 1. — P. 16048.
3. Quast, C., Pruesse, E., Yilmaz, P. *et al*. The SILVA ribosomal RNA gene database project: improved data processing and web-based tool. Nucleic Acids Research. — 2013. — Vol. 41. — P. 590-596.
4. Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A. *et al*. Uncovering earth`s virome. Nature. – 2016. – Vol. 536. – P. 425–430.
5. Hanson, N.W., Konwar, K.M., Wu, S.J., Hallam, S.J. Introduction to the analysis of environmental sequence information using metapathways. Comp. Meth. Next Gener. Sequenc. Data Analysis. – 2016. – P. 25–56.
6. Iverson, V., Morris, P. M., Frazar, C. D. Untangling genomes from metagenomes: Revealing an uncultured class of marine Euryarchaeota. Science. — 2012. — Vol. 335. — P. 587-590.
7. Bashenkhaeva, M. V., Zakharova, Y. R. Petrova, D. P. Sub-Ice microalgal and bacterial communitie in freshwater Lake Baikal, Russia. Environmental Microbiology. — Vol. 70, No. 3. — P. 751-765.
8. Mikhailov, S, Zakharov, Y. R., Galachyants, Y. P. *et al*. On the uniformity in taxonomic composition of bacterial communities in the photic layer of the three basins of lake Baikal, which differ in composition and abundance of spring phytoplankton. Reports of Academy of Sciences. — 2015. — Vol. 465, No. 5. — P. 620-626.

9. Gladkikh, A. C., Kalyuzhnaya, O. V., Belykh, O. I., Ahn, T. S., Parfenova, V. V. Analysis of the bacterial community of two endemic sponges from lake Baikal. Microbiology. — 2014. — Vol. 83, No. 6. — P. 682-693.
10. Zemskaya, T. I, Lomakina, A. V., Mamaeva, E. V. *et al*. Bacterial communities in sediments of Lake Baikal from areas with oil and gas discharge. Aquatic Microbial Ecology. — 2015. — Vol. 75. — P. 95-109.
11. Bukin, S. V., Pavlova, O. N., Manakov, A. Y. *et al*. The ability of microbial community of Lake Baikal bottom sediments associated with gas discharge to carry out the transformation of organic matter under thermobaric conditions. Frontiers in microbiology. — 2016. — Vol. 7. — P. 690.
12. Schloss, P. D. et al. Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology. — 2009. — Vol. 75 (No. 23). — P. 7537-7541.
13. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. — 2010. — Vol. 26, No. 19. — P. 2460-2461.
14. Johnston, W.M., Hanna, J.R.P., Millar, R.J. Advances in Dataflow Programming Languages. ACM Computing Surveys. – 2004. – Vol. 36. – P. 1–34.

Optical character recognition functionality introduction in mobile application for car diary

Ioannis Patias[1]

[1] University of Sofia "St. Kliment Ohridski", Faculty of mathematics and informatics,
5 James Bourchier blvd. (room 111)
BG-1164 Sofia, Bulgaria
patias@fmi.uni-sofia.bg

Abstract. The purpose of the paper is to develop a mobile Android application - "Car Log" that gives to users the ability to track all the costs for a vehicle and the ability to add fuel cost data by taking a photo of the cash receipt from the respective gas station where the charging was performed. Optical character recognition (OCR) is the conversion of images of typed, handwritten or printed text into machine-encoded text. Once we have the text machine-encoded we can further use it in machine processes, like translation, or extracted, meaning text-to-speech transformed, helping people in simple everyday tasks. Users of the application will be able to enter other completely different costs grouped into categories and other charges. Car Log application quickly and easily can visualize, edit and add different costs for a car. It also supports the ability to add multiple profiles, by entering data for all cars in a single family, for example, or a small business. The test results are positive thus we intend to further develop a cloud ready application.

Keywords: Optical character recognition, mobile application, car diary.

1. Introduction

The main focus of the application is in recognizing text in a photo of a cash receipt. Optical Character Recognition (OCR) is a technology for converting handwritten or printed text into machine-encoded text. It is widely used to file data from printed-paper, including personal documents, invoices, bank statements, computer prints, etc. This method allows us printed texts to be edited with a text editor, stored more compactly, displayed online, and used in computer programs such as automated translation or text-to-speech conversion.

OCR is applied in many applications. The main reason for the development of such applications is the weaknesses of the paper carrier as a means of storing information. The more paper is growing, the more productivity falls. It is becoming increasingly difficult for information to be found, and its use becomes slow and

inefficient. Paper information cannot be edited at the moment, as it would be possible to work with a text file on a computer. All these factors lead to the introduction of OCR, which allows the retrieval of obsolete, expensive and hard-to-handle paper and the fitting of text able formats that can be easily found, used and edited.

The process from the end-user's point of view is very simplified, as the user only scans the document, by taking a photo. Car Log app loads the scanned file into the OCR platform, and saves the recognized file in the selected format. We will describe in details the process latter.

Looking in Google Play Store [1] we may find some applications with similar functionalities. We compared some of them (Fuel Manager, Fuel Buddy - Car mileage Log, and Car Diary) to our application (Car Log). The apps are compared to some of the more important features of similar type of applications, such as: support to more than one vehicle, addition of charges and other type of expenditures, calculation of average cost and related information, number of charges and other similar features affecting fuel consumption, simple and accessible interface, as well as the innovative functionality in the Car Log, namely, OCR for the conversion of the information from a photo of the charge receipt. The comparison results are grouped in Table 1.

Table 1. Comparison of similar apps functionalities.

	Fuel Manager	Fuel Buddy - Car mileage Log	Car Diary	Car Log
support to more than one vehicle	-	+	-	+
addition of charges and other type of expenditures	+	+	+	+
calculation of average cost and related information	+	+	-	+
simple and accessible interface	-	-	+	+
OCR for the conversion of the information from a photo of the	-	-	-	+

charge receipt	

The main contribution of the paper is to show an easy way to build Android [2] mobile applications with OCR functionality.

2. Car Log application description

For the purpose of the demonstration, the Car Log application will be described in details, following the structure:

2.1. Description of the system model

The application is based on triple-layer architecture. It is divided into three linked, independent layers:

- Presentation layer - represents the user interface for interaction with the user. Performs the actions of displaying and collecting information;
- Business layer - represents the business logic of the application. Performs actions, related to data processing, reading and recording in the database;
- Data storage layer - represents the database in which the information in the application is stored.

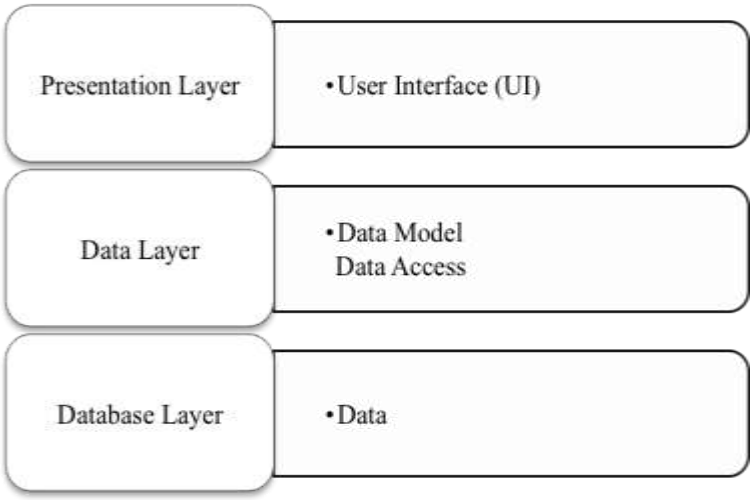


Figure 1: Triple-layer architecture application

2.2. Description of the database and the application

For the needs of the "Car Log" application, SQLite [3] is appropriate because of its simple interface, speed and flexibility. The complete SQL database containing tables, indexes, triggers and views is represented by a single file on the disk. The designed database consists of 4 tables. In every table is stored the data for the basic functional units of the application. The relations and dependencies between the objects, required for the proper operation of the application, are defined in advance in the following ER diagram (Figure 2).

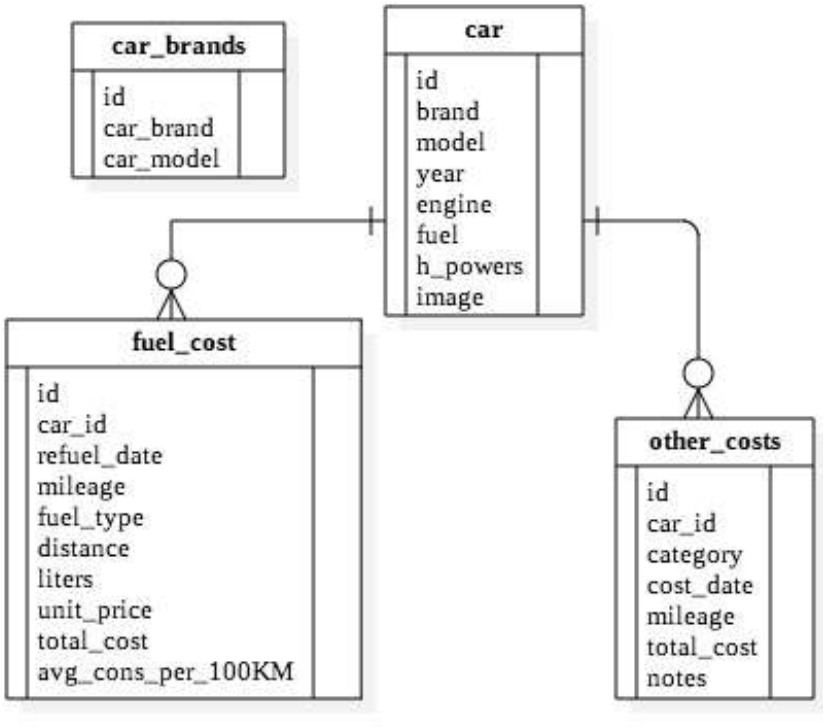


Figure 2: ER diagram for the database used.

The Car table contains the car profiles. It contains the following fields:

- id - a unique field for a profile
- brand – brand of the car
- model – vehicle model
- year – year of vehicle production
- engine – information about the vehicle engine
- fuel – information about the vehicle fuel type
- h_powers – vehicle's power
- image - a path that points to a file located in the file system representing a car image

The fuel_cost table contains information about the entered charging for a vehicle, which is written in the following fields:

- id – unique field for the given load
- car_id – an external key pointing to a record in a car table
- refuel_date – represents the date of loading
- mileage – mileage of the car
- fuel_type – type of fuel loaded
- distance – distance traveled since the last load
- liters – amount of fuel in liters
- unit_price – unit price per liter of fuel
- total_cost – total charge for the entire load
- average_cons_per_100_km – average charge for the specific load

The other_costs table contains information about other costs incurred for a given car, which is listed in the following fields:

- id – a unique field for that extra cost
- car_id – an external key pointing to a record in a car table
- category – consumption category
- cost_date – represents the expense date
- mileage – mileage of the car
- total_cost – cost of the expense
- notes – spending notes

Table car_brands is a table with only three fields. Vehicle and car model records have been entered in advance. The car_brand and car_model fields are TEXT type, car_brand cannot be NULL, and car_model can because the table is used in the business layer and there is an option to add a brand-only entry.

2.3. Implementation of the OCR functionality

OCR in the Car Log application is implemented through a server that accepts as input parameter the image in which the text is searched and returns a text file containing the recognized text to the image. The server used is the Abbyy Cloud OCR SDK [4]. This is a service, which is platform independent, due to the fact that it is accessible via the Web API and does not work on a device. The following activities diagram (Figure 3) describes the process.

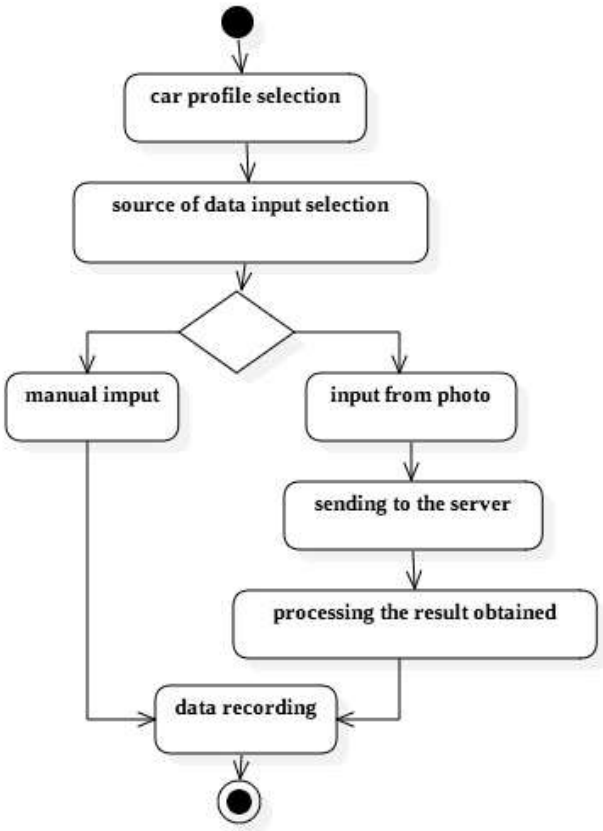


Figure 3: Activities diagram of the OCR functionality.

Of course there are various templates for receipts layout from on the market, and the application needs to take the layout differentiations into consideration. For the purpose the steps described in following activities diagram (Figure 4) are used.

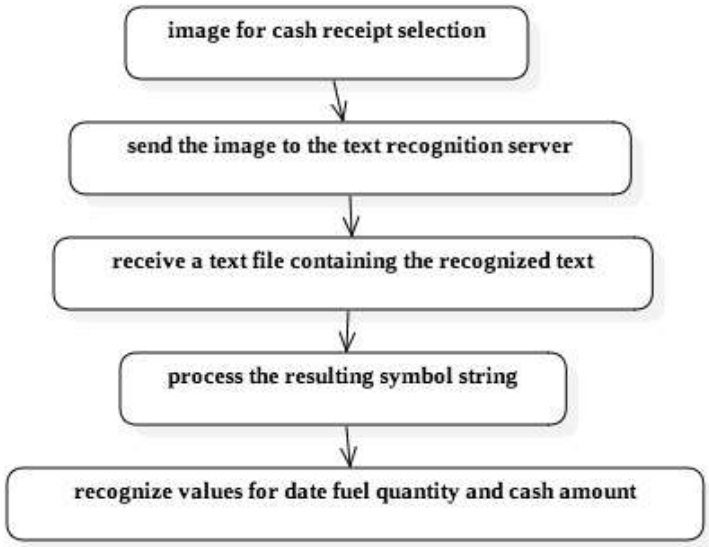


Figure 4: Steps for various templates for receipts layout recognition.

3. Test results

The conducted experiments and tests confirm that the Car Log application fulfills the defined goals and objectives. The application successfully maintains a database that stores fuel consumption information entered by the user, as well as other types of costs. The car's diary successfully copied and recognized text into cash receipts, from photos from the petrol stations.

The application was also subjected to a serious test related to the location of the camera in relation to the cash register. The following tests were performed:

1) photo of the cash receipt at a corner – in this case with all the receipts provided the result was the same, namely, the application did not cope with the recognition of the symbols and words in the photos and thus returned a failure recognition message,

2) long distance photo – with these tests bit of contradictory information was obtained – some images did not pass successfully, while in some others the application successfully read the information. Probably the cause of this difference is not in the distance, but rather in the image quality and more particularly the existence or the lack of focus,

3) photo in dark with the phone's flash – here the application has done also controversially. Some images were read perfectly, while in some others the only recognized information was the date, and

4) photo with lower light without the phone's flash – in this case there is no definite result – some images were successfully processed, some others not.

Test results show that the app performs fairly well, even when the photo of the receipt is not perfect. Only in the case of corner-snap tests the application categorically failed to process the photo, but in all other cases there was at least 50% success rate. When trying with normal photos taken in daylight and from a relatively close distance, the app did more than successfully.

4. Conclusions

By using the Car Log app, we can quickly and easily visualize, edit, add different costs for a car. The application supports the ability to add multiple profiles, meaning we can enter data for all cars of a family, or a small business. The interface is extremely accessible and intuitive for work. The application can be further developed in the future, and first of all, the database could not just be local, but could be uploaded to a cloud server. In this case, the application can be logged in, on any vehicle data, and thus, by multiple devices. Also, for the case we examined we use the technique of calling the server to do the OCR, but as further development we could incorporate the OCR in the application. The application is useful, thus further developments should be considered, especially related to development of a cloud ready application.

5. Acknowledgements

This paper is supported by the project ДН 02/9-17.12.2016.

6. References

1. https://play.google.com/store/search?q=car%20diary&c=apps (April 2017)
2. Bill Phillips, Chris Stewart, Kristin Marsicano, “Android Programming: The Big Nerd Ranch Guide” (3rd Edition) ISBN-13: 978-0134706054, ISBN-10: 0134706056 (2013)
3. https://developer.android.com/reference/android/package-summary.html (2017)
4. http://ocrsdk.com/documentation/apireference/processImage/ (2017)

Internet's Impact on Teaching

Olivera Iskrenovic-Momcilovic, Dragan Cvetkovic

University of Novi Sad, Faculty of Education, Podgoricka 4
25000 Sombor, Serbia
oljkaisk@yahoo.com, dcveles@gmail.com

Abstract. The Internet is an indispensable part of our everyday life. It provides the ability to spread knowledge and information among a large number of users who are spatially distant. The application of the Internet in teaching brings many advantages, but also disadvantages. The benefits are learning independent of time and space, editing capabilities, documentation and reuse of teaching content, access to other digital sources of information; As well as a new learning culture through the concept of team, cooperative learning. The disadvantages are reflected in the technical problems and the lack of acceptance of users, the lack of social indicators, where a phenomenon, such as social laziness, can aggravate virtual collaboration. Problems are also wasting valuable time sitting in front of the computer. Also, the content that is available to young people today, if they are violence, pornography, contact with anonymous users, suspicious virtual friends, especially those who represent themselves as peers.

Keywords: teaching, Internet, web search, e-mail, social networks

1. Introduction

The Internet has become the largest resource of information and the fastest way of communication. Teaching is based on communication and information exchange, so the Internet is ideal for teaching. It affects collaboration, motivation and (self) evaluation, as a medium suitable for connecting with the whole world. The application of the Internet in education can increase the quality and efficiency of teaching, which is increasingly characterized by interactivity. When an individual searches the Internet, he will come across a variety of useful information. In this way, it will be easier to form a more complete and objective image on a particular subject matter.

The Internet overcomes spatial and time constraints, just as human thinking does. It is necessary to use all the opportunities provided by the Internet in education. It is a powerful tool for learning and research, if an individual knows that he is using it the right way. Today, the biggest problem is the ability of an individual to implement information and communication technologies. An important task of teachers is to teach students how to use the Internet in order to enjoy the benefits offered by and abandon abuses.

In a classroom full of students who spend a lot of time typing on their smartphones and (unconsciously) visiting the Internet, they can easily detect the impacts of information-communication technologies. Now, the student has a huge amount of information at your fingertips or is even "hand-in-hand". On the other hand, teachers should take advantage of the ability of the Internet and multimedia tools to better maintain their lessons, to give them a new, more flexible form that will benefit more students. In particular, the Internet and information and communication technologies affect every aspect of education.

2. Internet and teaching

a) Web search

The most accessible and most valuable Internet service for information transfer is *WWW (World Wide Web),* a simpler *Web*. It allows the site to display pages with a large number of information in the form of text, images, sound and animation. Most pages are linked to other sources and are referred to by users on their pages. The Web is based on hypertext technology. With hypertext, one document can be linked to others of different content, wherever they are remembered. This creates non-transparent sets of new information and the user can easily "lose" in the mass of information. Fortunately, there are web browsers, Internet browsing programs. Which are constantly improving in order to be simpler, more accessible and more efficient for "surfing" and transferring information. Browser is a tool designed to search information on the Web, most often by keyword setup. Based on the degree of similarity of these key words and pages, the bids ranked by the user are ranked. There are many browsers today, but the most famous are: *Google Chrome, Mozilla Firefox, Internet Explorer, Opera, Safari*, and so on (Fig.1).

Figure 1. Web browsers

Web search speeds up the discovery of new information, which can become future knowledge. This in any case contributes to increasing the quality of teaching. Teaching

enriched with Internet sources provides more learning opportunities than before. The Internet is a great inexhaustible resource for teachers. Teachers can find on the Internet suggestions and plans of lessons, expert suggestions, practical support, necessary literature and other teaching materials, which makes preparation for quality teaching easier. There is also a range of site-owned organizations that can be a useful resource to teacher centers. Thus, with the help of the Internet, a teacher can explore any subject, improve and prepare for the time or any other educational task. Some search engines are especially suitable for teachers and students who are interested in improving the quality of teaching. [2] Some of them are: *www.bigchalk.com, www.education-world.com, www.awesomelibrary.org, www.ask.com, www.refdesk.com ...* Special emphasis is given to *Google Schoolar*, a search engine for scientific literature From different areas and sources.

Due to the enormous amount of information that the student is exposed to, he should be able to collect them, evaluate their quality, filter them, synthesize and apply them. The teacher must be able to do the right thing and teach his students. Now, when it's no longer hard to find any information, it must become an instructor, which will help the student assess and "overwhelm" a huge amount of information in search of valid ones. [3] This will allow students to learn faster, to approach topics in more detail, to make homework more easily, after all, understand and feel how to search for knowledge. Of course, there is a risk that some students will only reach for the fun content that the Web offers or that easy access to information will provoke plagiarism. This should be prevented by a teacher, as a good and influential role model. It should be noted that young people are increasingly living and progressing in the digital environment and are able to access a wide range of resources and services. So connected they simultaneously do homework, talk on the phone, listen to music, search the Web and maintain online conversations. This is also supported by the fact that the use of information and communication technology from an early age still improves learning. [4]

b) E-mail

The Internet specifically contributes to user communication. The user decides whether to get in touch with a written or video message or even face-to-face contact with the person you call on the Internet. If a document is required to be transmitted or received, it does not pose a problem, as it is sent over the Internet in different forms.

Teachers and students around the world, noting the benefits of this kind of communication, began to exchange information with each other in a new way. With better communication, teaching content is more appropriately adopted, conversation and working together on tasks is easier, and knowledge is gained faster. [5] The possibilities of using the Internet in teaching continue to increase every day. His potential for improving education is enormous, but it is necessary to have awareness of possible misuse of this useful asset.

Nearly no person has access to the Internet, and there is no e-mail. Electronic mail is one of the first Internet services. Today, it is by far the widest, because it allows for quick exchange of user messages, which can be found anywhere in the world. Electronic letters arrive in the mailbox immediately, and can be read at any time. A part

from text messages, visual, audio and audiovisual records can be sent as an attachment. As the best providers of e-mail services are (Fig.2):

- *Gmail* - simple and organized to store information
- *Outlook* - former Hotmail, allows video chat and chat with friends,
- *Yahoo and Zoho Mail* - popular youth code

Figure 2. E-mail

Teaching and additional materials can be sent using e-mails [6]. Therefore, only two people can participate in the communication, which is suitable for individual work, as well as the mentor work of the teacher with the student. Then, private messages are sent, which can only be seen by interlocutors. Messages can be sent simultaneously to multiple recipients, groups. In this situation, e-mail lists are created, where people who are interested in certain content are selected. This makes the whole communication much simpler and the issue of informing associates (teachers and pupils or colleagues) is not a problem. [7] E-mail is a convenient way to e.g. One teacher informs his students about an event that concerns them or reminds them to do tasks or help with the delivery of literature. Students can help each other, learn together and fulfill all tasks. Communication via e-mail brings students into the real world and teaches them a bonton of written (and oral) speech that needs to be mastered and respected.

c) Social networks

"Study group" means gathering students in order to solve tasks together and learn. Today, this collaboration has largely transmitted "online" because it is more often communicated via social networks in such situations today. A social network is a type

of Internet service whose purpose is to connect users. It becomes the leading medium for Internet communication. [8] From the chat rooms as their first forms, they have improved and brought new opportunities. However, apart from the entertainment side, which first comes to mind, social networks can be used wisely and useful, which emphasizes their informative-educational function.

Figure 3. Social networks

Facebook was designed to interact with Harvard students, where he studied his inventor Mark Zuckerberg. However, today, Facebook, as a global social network, is used for various purposes. It provides quick information sharing, publishing activities on the "wall" of the group that all members can see, send pdf and Word files, share multimedia content, and applications such as video games and quizzes, which can be fun for students. *Twitter* serves to send quick, short messages (tweets), images and links, which are tagged and thematically sorted. *LinkedIn* is the largest business network that connects businesses, business partners, and educational institutions. It enables group membership, where various topics are discussed. *Google+*, *Flickr*, *Pinterest and Instagram* are also popular social networks that connect people and facilitate the sharing of diverse data.

In addition to strengthening interpersonal relationships, networking encourages participation in socially useful activities, cooperation with schoolmates or colleagues at work, work on group projects and sharing achievements. Access to new resources, especially human, is more accessible than ever, and such communities can find and analyze anything. As soon as a question is raised on the social network, the comments or responses will follow as soon as possible, and the individual can become a resource to others by engaging in a conversation. [9] Social networks have reduced the generation gap and have obviously improved the cooperation of teachers and students - it has become more practical and open.

In free time students and teachers use some social networks, which indicates that they are interested in this kind of communication. With the help of the Internet, it is easy to get in touch with colleagues, mentors, professional centers, materials sources, regardless of the physical distance. It is particularly useful that social media can track novelties in different fields, sharing resources and exchanging ideas about educational

content. It is unavoidable to introduce social networks in teaching, in order to make their opportunities even more useful.

In addition to numerous advantages, which are undeniable, there are also gaps in social networks. Using social networks, direct communication is neglected, and dictionary debris has been depleted, dominated by jargon and strangers. Although social networks are considered to promote bad spelling, it is the teacher who gives a good example and corrects students on the right path in the social networking world. [10]

In the world of social networks it is bad that, due to hacker attacks, abuse of shared data, especially personal, can occur. Publications and statuses on social networks remind of former secret diaries of young people, just as they are public! Also, the virtual space corresponds to introvert people, but there is a danger that network connectivity is perceived as close, although it is not. [11] Social networks divert too much attention, so a measure is needed to focus our attention on quality conclusions and knowledge.

3. Advantages and Disadvantages of Internet Impact on Teaching

The Internet brings many benefits to teaching and removes some of the shortcomings of the traditional approach. With the Internet, students can search content from many sources, and share them with different devices, with others in real time. [12] In this way, the Internet becomes complementary to standard textbooks and allows for the creation of a wider image, as well as deepening of the topics that students are particularly interested in. It is very important that one can learn at any time, independently determines the pace of work, the speed of the search, and the style of adoption of the material, so that the individual progresses in accordance with his personal abilities and aspirations, which are characteristics of individualization. The flexibility of learning, as well as the dynamics of the overall activity, has increased. This gives students a sense of freedom, so they can easily explore, learn and check their knowledge without a press. Communication over the Internet and distance learning are suitable for people with disabilities or those living in inaccessible rural areas, because they give them more room, a more comfortable and safe atmosphere (for example, if they go home from home), increases the participation and availability of training materials. Also, adequate monitoring and fast feedback are there to support the development of persistence, persistence and precision, while students are self-seeking for knowledge. The Internet is also "saving" for shy, sensitive and scary students, because it reduces the possibility of frustration due to failure.

The Internet in teaching is especially useful to gifted children, who are always ready for more thorough and deeper knowledge. It allows the content to be adapted to the greatest extent by the majority of students. But Internet teaching will not only be used as a search engine for finding new facts, but also as a link between different schools, students and teachers, for their cooperation and common progress, and may be the initiator of mobility. It is possible to further improve, attend online training and courses, which will bring a solid theory for practical skills. [13]

Therefore, the Internet increases the potential of learners for learning, but also the potential of knowledge transferring teachers, opening up both access to a huge amount of information to both. Now the teacher can upload documents, e-books, multimedia content or quizzes and knowledge tests, which can automatically reach students. Tasks can also be published online, while ratings will be viewed in a virtual space, such as school electronic diaries. In this way, communication, interaction and cooperation between students, teachers, and parents is facilitated. The Internet offers options for sending emails, chatting, compiling documents, blogging, sharing information, social networking and more!

However, it has also been shown that the Internet is a two-blade sword because it also has many negative effects. All its users must be familiar with the dangers that the Internet has to avoid in order to avoid potential pitfalls. One of the most current questions is whether the Internet helps or defrauds a child's social life. On the one hand we are witnessing that the Internet is the cause of bad social development of children because the computer has become a means by which almost all children know how to communicate. There is a problem that the usual activities that concern the social development of children endanger. The basic objection is the lack of real interpersonal relationships, primarily face-to-face face-to-face communication with teachers and peers. Although the Internet essentially connects, it is like losing the effects of collaborative work, because the student intensively "communicates" with online content, and only if necessary with his associates or mentors, thereby reducing socialization.

Since everything is "click-through", the need for memory and imagination is reduced. As much as it can be developed, the development of creativity, divergent and critical thinking can be neglected too much. It depends on the user whether it will permit the risk of overwhelming the Internet-only knowledge-based knowledge-based knowledge that will not lead to the insight of the situation of situations and the deepening of knowledge or the development of the abilities of creative abstract thinking and creative learning. The pupil can easily become passive if there is no strong motivation for work, developed will, control, and self-control. With all this patience patience. Children who grow up in the Internet era are waiting for "instant" satisfaction. They can easily find everything they want online with minimal effort, and everything in real life does not go the wrong way. Psychological dependence can also occur, when the virtual "society" of the Internet becomes a refuge from genuine problems.

A special item is an increasing problem of data abuse from the Internet. First of all, data shared on the Internet is poorly protected and extremely inhumane can be exploited, and, secondly, students themselves are increasingly resorting to cheating. Ready-made papers or test responses can be found on the Internet wherever the issue is concerned. Knowing that they are easily accessible on the Internet, many will be tempted to do so, and they will become plagiarists and fraudsters, which will not be important knowledge and value, but pseudo vocation and "finding". Unfortunately, it is often difficult to monitor, monitor, and prevent this kind of mischief. [14]

It is obvious that hours before the screen and a rare change in position have negative health effects. Namely, due to the long stay in the sitting position, the body is held down, and weak and visionless. Surfing on the Internet will most likely bring useful

information of a different type, but this type of "activity" can drastically interfere with physical development, especially children and young people (who are also the most common Internet users). [15] If someone keeps on the Internet for a long time - working on your brain, it would be most preferable to spend at least half the time in any physical activity - working on your body.

However, one needs to be rational and respect both the advantages and disadvantages. The use of the Internet in teaching is one of the good ways to innovate and dynamise the teaching process, but it should not be overstated. Namely, the Internet has proved to be the best means of completing, enriching, deepening and teaching content or independent and group work on specific projects or research. It is in these fields that he and his abilities need to be used to the maximum. Today, the Internet is widely applied at all levels of education: from pre-school, school, university education, to adult education, and to extra-curricular education and activities. But it must be kept in mind that its teaching advantages are manifested only in the flexible approach of learners to learning tasks, which often does not allow a strictly defined duration of time.

4. Conclusion

The use of the Internet in teaching provides a number of benefits. Learning with the help of the Internet is individualized and autodidactic, diversity, interestingness, modernity, better motivation, communicativeness, transparency, repetition and repetition to a complete understanding and better memory of the content are introduced. The student is testing himself and adjusting the pace of learning with his abilities. The most basic is the knowledge that has come about through its own engagement and conclusion, which certainly enables the Internet to provide a multitude of different information.

However, education is lagging far behind in the application of information and communication technology. There are many reasons for this, beginning with the lack of hardware, software, to conservative teaching staff and rigid education development strategies. This, however, relates to the formal presence of the Internet in the classroom, since the Internet is certainly included in the life of students and teachers who, with the help of him, solve various educational tasks on a daily basis. The Internet is an exciting resource for everyone who is related to teaching. It is a tool that allows students and teachers to connect more easily and discuss more often by highlighting various topics.

The Internet is perhaps the most powerful tool at the moment, which can be applied in class because its mode of operation supports the idea of self-improvement and lifelong education. It allows more people to get involved in the teaching process and to become more and better educated (of course, if they know that they have to invest their time and effort, as well as respect for others' advocacy), with less costs and a greater choice of topics and subjects . But, in order for the Internet to really deliver well, it's necessary to get acquainted with its wise use. The Internet is basically intended for education, and only later it is followed by entertainment, which many often overlook or fail to see.

Interest in this kind of work is spreading rapidly around the world and certainly its opportunities should be used. But, in addition to the whole modernization, one must not forget that people create people and that the role of teachers, his human character and work are the most important condition for the development of students. Also, it is not possible simultaneously to make substantive and organizational changes, which imply scientific and technological and social development, while retaining the traditional traditional organization of work. This is an obstacle, which restricts the dynamic development of education. It is necessary to build such a school, such an educational system, in which the teaching will be in the function of comprehensive development of young people and social progress in general. We must make sure that on this path the Internet is also a useful associate, which can really be.

References

1. Halverson R., Annette Smith, A.: How new technologies have (and have not) changed teaching and learning in schools, Journal of Computing in Teacher Education, Vol. 26, No. 2, 49-54. (2009–10)
2. Purcell K., Buchanan, A. J., Friedrich, L.: How teachers are using technology at home and in their classrooms, Pew Research Center's Internet & American Life Project, Washington, USA. (2013)
3. Taylor, A.: A study of the information search behaviour of the millennial generation, Information Research, Vol. 17, No. 1, 1-10. (2012)
4. Drajden G., Vos DŽ.: Revolucija u učenju – Kako promeniti način na koji svet uči. Timograf, Beograd, Srbija. (2004)
5. Buabeng-Andoh, C.: Factors influencing teachers' adoption and integration of information and communication technology into teaching: A review of the literature, International Journal of Education and Development using Information and Communication Technology, Vol. 8, No. 1, 136-155. (2012)
6. Hassett, J. M., Spuches, C. M., Webster, S. P.: Using electronic mail for teaching and learning, To Improve the Academy, Vol. 14, 221-237. (1995)
7. Brändström, C.: Using the Internet in education – strengths and weaknesses, University of Gavle, Gavle, Sweden. (2011)
8. Vasilijević, D.: Društvene mreže u obrazovanju: stavovi i očekivanja studenata, Teme, g. XI, br. 4, 1241-1258. (2016)
9. Glušac, D. Elektronsko učenje, Tehnički fakultet Mihajlo Pupin, Zrenjanin. (2012)
10. Šmakić, K.: Društvene mreže i mladi: utjecaj opcije 'like' na kreativno ponašanje mladih, In medias ras - casopis filozofije medija, Vol. 5, No. 9, 1333 – 1338. (2016)
11. Putnik, N., Babić, L., Kordić, B.: Socio-psihokoški i bezbedonosni rizici naručavanja privatnosti na društvenim mrežama, Međunarodna naučna konferencija Sinteza 2014, Univerzitet Singidunum, Beograd, 683-686. (2014)
12. Raut, V., Patil, P.: Use of social media in education: Positive and negative impact on the students, International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 4, No. 1, 281 – 285. (2016)
13. Eluwole, O. T., Udoh, N. S., Ojo, O. D.: The impact of Internet on African education and culture, International Journal of Business, Humanities and Technology, Vol. 4, No. 3, 69-77. (2014)

14. Gómez-Espinosa, M., Francisco, V., Moreno-Ger, P.; The Impact of activity design in Internet plagiarism in higher education, Comunicar, Vol. 24, No. 48, 39-48. (2016)
15. Antić, J., Zelić, M., Jančić, G.; Društvene mreže, životna sredina i zdravlje, 11. Međunarosni simpozijum Infoteh, Jahorina, Bosna i Hercegovina, Vol. 11, 725 – 729. (2012)

International Conference Paper Tracking System

Zoltan Kazi, and Ljubica Kazi

University of Novi Sad, Technical faculty "Mihajlo Pupin"
23000 Zrenjanin, Serbia
{zoltan.kazi, ljubica.kazi}@gmail.com

Abstract. This paper presents a paper tracking system developed at Technical faculty "Mihajlo Pupin" in Zrenjanin. It was created and used at Ecology of Urban Areas international conference in 2016. This system is a web application that handles data about conference, participants and their papers in relational database. It is an integral part of the faculty web site and enables paper author and user registration, user administration, paper submission, paper reviewing, conference registration and generating various reports and statistics.

Keywords: paper tracking system, international conference, web application.

1. Introduction

At this moment, there are various solutions for papers submission and their tracking. These solutions are mostly used in international scientific journals, while at conferences and symposiums the situation is quite different. Communication with editors, scientific board members or reviewers is mostly done with e-mail. Contemporary scientific conferences use online systems to handle paper submissions and reviews [1]. Technical Faculty "Mihajlo Pupin" in Zrenjanin, before this software, did not have such a system. Due to the growing number of papers and authors who apply for participation at six international conferences with very short deadlines, large number of messages that must be sent to authors and reviewers, it is turned out that it is necessary to introduce such a system. Introduction of these systems has significantly facilitated the administration, submission and review process compared to traditional paper-based ones [1].

This paper introduces a software system for paper submissions at technical faculty in Zrenjanin. The rest of the paper is structured as follows: we present existing solutions for paper tracking and submission systems in Section 2. In Section 3, we describe system requirements, software functions, database design, and web application programming important parts. In Section 4, we present the implemented system, web application description and generated reports. Conclusion is in Section 5, while Section 6 lists the paper references.

2. Existing Paper Tracking Systems

This section presents some of existing solutions for paper tracking that are used in recent years at MIPRO international conference, Easy Chair, Cite Seer, Springer and Elsevier groups for publishing numerous scientific journals and other publications.

Mipro is an international convention organized by Croatian Society for Information and Communication Technology, Electronics and Microelectronics - MIPRO. Seat is in Croatia, in the city of Rijeka. This convention is held for 40 years in Opatija tourist place, near Rijeka in Croatia. This convention has been using a system for papers tracking for more than ten years. This system is presented on Fig. 1 and it enables: user registration, user login, paper submission for more ten sections, paper status tracking, paper reviewing, and administration. Users can withdraw their work, while reviewers can accept or reject paper reviewing without any explanation. [2]

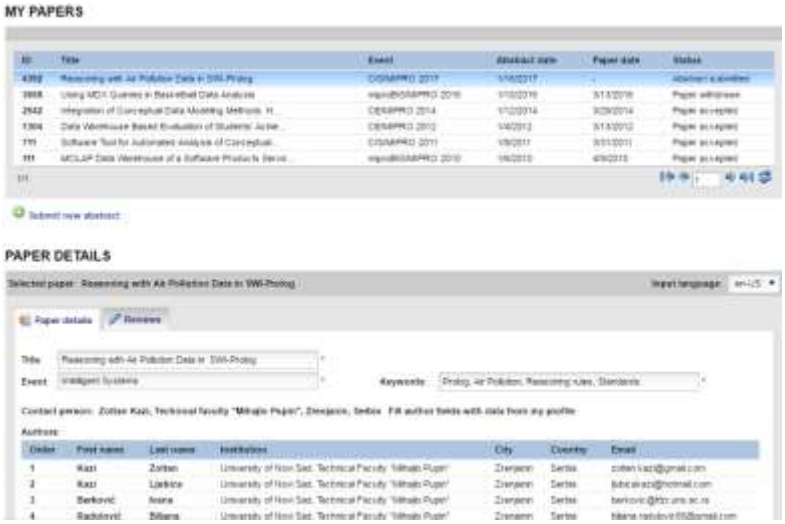


Fig. 1. MIPRO convention paper tracking system, part for authors

Another system is Easy Chair [3]. This system was developed in the year 2002 by Andrei Voronkov, a professor from the University of Manchester [1]. This system is used by several hundred conferences and conventions. In Serbia, for example, this software is used on information society and technology – ICIST which is held for seven years at Kopaonik ski center in Serbia. This system supports different kinds of licenses with their features and prices. There are restrictions on the number of free licenses that can be issued to a single conference. When users initialize a new conference, it will be automatically given the free license. Non-free license must be paid. Easy chair features are: hosting on a high-performance server, data backup and replication, encrypted (secure) access only, helpdesk and technical support, import of configuration from other conferences, branding, support for anonymous submissions (double-blind reviewing), customized submission types, author notification, instruction for authors, custom fields in the submission form, multiple file uploads, audio, video and zip file uploads. Program committee and reviewing part has following features: event log, watch list, online discussion of submissions, flexible access models for reviewers, sub reviewers, customized decisions (acceptance types), customizable evaluation criteria, statistics, custom review forms, multi-track support etc.

CiteSeer is a digital library that was developed for automatic tracking of scientific literature that is relevant to a user's research interests. Uses can examine its database of publications to determine whether any new papers are related to the user's interests, and then the user can be alerted by e-mail or whenever they next use CiteSeer's web based user interface. A heterogeneous relatedness measure is used to identify new related documents. Also, citation links can be monitored to discover new citations to existing papers.

Springer is a global publisher group with 175 years long tradition that enables scientific, technical and medical portfolio for researchers in academia, scientific institutions and corporate departments. Springer basic services for paper tracking are: managing users and their accounts, electronic paper submission, papers reviewing and acceptance handling, managing copyright, author help desk, etc. [4]

Elsevier is with Springer the global leader in science, technical and health publishing founded in 1880. More than 25000 academic and government institutions around the world use Elsevier products. It provides information analytics solutions and digital tools in different areas. Elsevier has tools and resources for reviewers, services for electronic paper submission and tracking, user registration, creating authors own personalized home page, etc. [5]

Research study into the status of online submission and peer-review systems, the perceptions of these by authors, referees and editors, and the impact of their introduction on journals is presented in report [6].

3. System Requirements and Design

The requirements set by the organizing and scientific board members were:

- simple login and user registration with less input data;
- possibility to select exactly two reviewers for every paper;
- reviewing papers with word template that was used until 2011;
- automatic e-mail sending after every step that is made by conference secretary;
- creating various reports that could be done and finished in electronic form or has to be printed;
- new functions must be incorporated into faculty web site and pages for conference.

Before building such a system it were created simple models that describes user software functions through use case model and database was projected first at logical level with conceptual data model, and then on physical level with physical data model.

3.1. Software Functions

Software functions are projected from requirements list that are agreed at organizing committee and defined on the features of the previous described systems [2],[3]. They were realized in use case model presented with diagram on Fig. 2. There are three type of users: authors of the papers, paper reviewers, and system administrator/conference secretary. Authors can fill user registration form to become a system user. After that users are able to submit a paper, and to add coauthors for papers. Finally, they can fill out the registration form for conference attending. Reviewers' functions are: user registration, paper submission, managing coauthors for papers, adding reviews, and filling out the registration form for conference attending. Administrator has more functions than other users: managing users, paper submission, reviewer selection, creating various reports, and creating conference statistics.

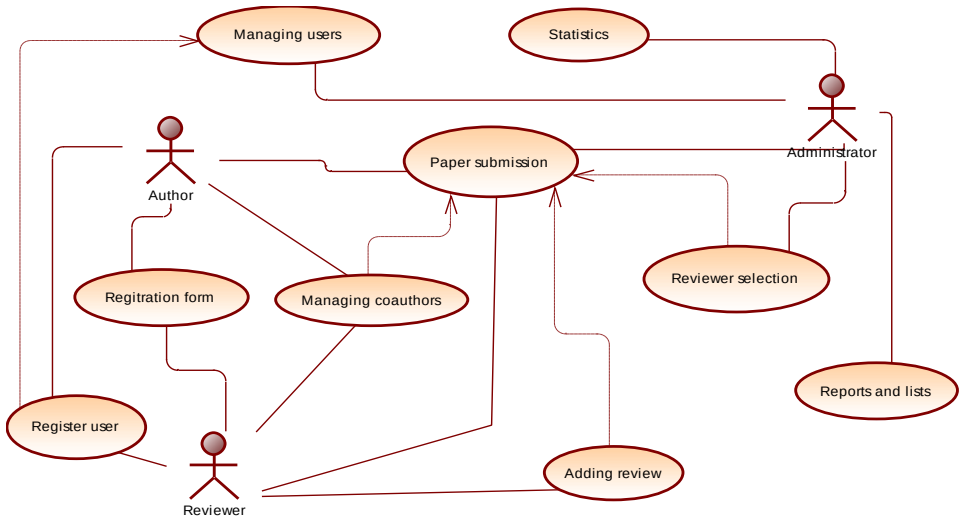


Fig. 2. Paper tracking system software functions

3.2. Database Design

Relational database was created from physical data model that was first designed at logical level in conceptual data model. Entities present at conceptual model are: user, paper, author, conference, thematic field, paper reviewers, and registration.

User and author attributes are: name, surname, organization, address, city, country, password, username and email. Every user can be only one author that can submit several papers. Author could have several coauthors that are authors also, but they are not registered as system users.

Every paper belongs to exactly one thematic field, while one thematic field has several papers, but it is possible to have fields without papers. Thematic field entity attributes are: id and field name. Paper attributes are: id, title, abstract, abstract accepting, key words, paper file, corrected file, reviews, paper sending date, paper acceptation date, and paper status. Every paper must be connected to one author and one conference. Conference attributes are: id, institution that is the organizer, city, country, and the year.

Papers and users are connected with paper reviewer entity. One paper has many reviewers, while one user can be reviewer for several papers. Restriction that one paper has two reviewers is done with reviewer1 and reviwer2 attributes that belongs to this entity.

Finally there is a registration entity with numerous attributes fee paying: id, way of presentation, pin, contact phone, registration number, code, bank name, authorized person, evidence number or vat tax payer, phone number, contact person, invoice, and registration date. One paper is minimum connected with zero registrations and maximum with several. One registration belongs to only one paper.

3.3. Web Application Programming

Database was implemented in MySQL database management system with SQL statements: create database, create table and alter table for reference and restriction handling. Web site communication with database is realized in PHP programming language. User interface and design of pages were done with HTML and CSS, while JavaScript was used for menus implementation. Reports and pages for statistics are simple web pages with white background, while data was retrieved from database with PHP commands.

The whole web site for Urbaneco conference has around one hundred PHP files with commands that realizes user functions defined with Use Case diagram. For every user that is successfully logged into the system it is created a session, so unregistered users are automatically redirected to the login page. Pages for user registration, paper submission, coauthors adding, paper reviewing, and registration for conference attendance are done with basic web forms. Input data was sent to another page with Post method. Inserting, updating and deleting data from database is realized with PHP Mysql_query builtin procedure:

mysql_query($Sql, $db)

where: $Sql variable can be Insert, Update or Delete SQL statements, while $db is database connection.

Sending e-mail messages was done with PHP Mail procedure, for example:

mail($email,"Urbaneco paper submission system",$msg,"From: urbaneco@tfzr.uns.ac.rs")

where parameter $email is authors' e-mail address, and $msg is massage text.

4. System Description

The whole system has more than one hundred PHP files. They all are integrated with existing static web site created in HTML with CSS and menus created with JavaScript language.

4.1. Web Application

Web application user interface is shown on Fig. 3. Users must first register, fill the registration form and then enter the user name and password. The only built-in database account is the administrator and it should not be registered. This account maintains the entire system and can change all data and determine the group for each user. All users are initially authors of works and papers with the status called "user". The administrator can, after registering, determine whether the new user will be a reviewer, administrator, or a regular user.

Fig. 3. Urbaneco conference paper tracking system – a part for paper submission

After successful registration and login, users are able to submit a paper, edit personal data and affiliation. After sending the work, authors can manage the coauthors of the paper, while the administrator determines exactly two reviewers. Through the system the users can have an insight into the current state of the submitted papers (Fig. 4).

Fig. 4. Urbaneco conference paper tracking system – a part for paper information and access

The reviewers receive mail and a notification on the necessary process. After completion of the review, the reviewer is has to login to the system, to find a paper and upload the review file

that has standard template form created in Microsoft Word. Work can be accepted without corrections, can be withdrawn (abandoned), and can be requested changes for final accepting. All changes must be uploaded again in a corrected file. Finally, after review process finalization, if the paper is accepted, system sends an e-mail message to the corresponding author who can fill out the registration form for conference attending.

Reviewers can be paper authors, also, so they can upload paper and manage coauthors. Other items on the reviewer menu are: statistic of all reviewers and their work, and downloading the template for the review process.

The key role in using the system is in the work of an administrator who can be a person in charge or a secretary of the conference. As it is mentioned the system administrator has more functions than all other users: adding data about conference, entering thematic fields, managing users, editing user data, paper submission correction, reviewer selection, creating various reports, and creating conference statistics.

4.2. Reports

Authors and reviewers cannot generate reports. Those activities are done by the secretary, i.e. administrator. Reports that could generate this web application are: list of all submitted papers with status (Fig. 5), list of accepted papers, list of participants, registration form, and statistic according to paper number and number of papers by country.

5. International Conference URBANECO
Palic, Serbia, 2016

LIST OF PAPERS

No.	Title	Authors	Country	Status
1.	ACCIDENTAL EVENTS OF PRIMARY SHORT-CIRCUIT OF ELECTRICAL INSTALLATIONS »Tr« FORENSIC ANALYSIS	Vjekoslav Sajfert	Serbia	abandoned
2.	ACCIDENTAL EVENTS OF PRIMARY SHORT-CIRCUIT OF ELECTRICAL INSTALLATIONS »Tr« FORENSIC ANALYSIS	Vojkan Zoric	Serbia	abandoned
3.	ANALYSIS OF THE IMPACT OF TRAFFIC DENSITY ON AIR QUALITY AT TOLL BOOTH NAIS ON HIGHWAY E75	Nikola Misic	Serbia	abandoned
4.	APPLICATION OF NATURAL INORGANIC SORBENT (PEMZA) FOR REMOVAL OF Cr(VI) IONS FROM WATER RESOURCES	KIRIL LISICHKOV	R. Macedonia	abandoned
5.	APPLICATION OF COMMUNAL SEWAGE SLUDGE TO ACIDIC SOIL INFLUENCES THE PLANT GROWTH AND DRY MATTER OF PHASEOLUS VULGARIS	Hamed Abdorham	Libya	abandoned

Fig. 5. Urbaneco conference paper tracking system – report for list of papers

All reports can be recorded electronically, e.g. In PDF format or print on paper. One of the basic advantages of this system is in the ability to create various reports that are obtained extremely quickly, if all the data of the authors, papers and reviews are entered through the system.

5. Conclusion

Use of a paper tracking system and electronic paper submission reduces the review process and reviewing time, which surely improves the work of the organizing committee. Existing solutions provide many automatization options and could be applied within any conference, but some conferences require specific business rules and data.

This paper presents created paper-tracking system for the need of specific conference "URBANECO". During creating this system, all the requirements, that were defined, were implemented. The design and implementation of the system was performed within short time period, for the conference to be held in 2016. Surely, there are many improvements to be made. Options like automatic e-mail sending after every step (previously made by conference secretary) and quick and easy report generation would significantly improve the quality of the conference organizer's work. This system could be upgraded to be general conference-organizing software support system for all international conferences that are organized at "Mihajlo Pupin" faculty, University of Novi Sad.

6. References

1. S-W. Lo, R.C.W., Phan, B-M. Goi. On the Security of a Popular Web Submission and Review Software (WSaR) for Cryptology Conferences, Infoscience, EPFL's scientific publications, https://infoscience.epfl.ch/record/111550/files/LPG07. (2008)
2. MIPRO International Convention system, available on: http://www.mipro.hr. (2017)
3. EasyChair conference system, https://easychair.org/licenses.cg. (2017)
4. Springer submission system, available on: https://www.elsevier.com/authors/journal-authors/submit-your-paper. (2017)
5. Elsevier submission system, available on: https://www.springer.com/gp/authors-editors/journal-author/journal-author-helpdesk/submission/1302#c1262. (2017)
6. Mark Ware, Online submission and peer-review systems, Learned Publishing, Vol. 18, 245–250Volume 18, Issue 4 October 2005, Pages 245–250, (2005).

Preschool Web Portal Development

Ljubica Kazi[1], Zoltan Kazi[1], and Tatjana Lojović[2]

[1] University of Novi Sad, Technical faculty "Mihajlo Pupin" Zrenjanin,
Djure Djakovica bb, 23000 Zrenjanin, Serbia
ljubica.kazi@gmail.com, zoltan.kazi@gmail.com
[2] Preschool Institution Zrenjanin, Karadjiceva 3a,
23000 Zrenjanin, Serbia
pomdir@predskolskazr.edu.rs

Abstract. Preschools are important part of formal educational system. Information technologies are included in preschools in various ways. This paper aims to analyze the requirements for preschool web portal development, compare existing solutions, as well as to represent an example of a developed solution. The developed solution is installed at Preschool institution Zrenjanin, Serbia and it is in use since January 2015. Plans for further improvements of the presented example web portal are also described in this paper.

Keywords: preschool, web portal, software development, PHP.

1. Introduction

Preschools refer to the first stage of formal education. There are many different terms related to preschool – kindergarten, daycare service, nursery school etc. Institutionalization of care and education of children is supported for different age interval in different countries and it is related to the primary school starting age. In some west Europe and English-speaking countries, preschools cover age of 2-5 and at the age of 5 children enter primary schools. In Serbia, preschool offer services for age 1-6, while at the age of 7 children enter primary schools. In Serbia, preschools is organized as government run and privately operated. Government-managed preschools are partially financed by local municipality funds (organizationally – all preschools are managed and supplied from centralized regional preschool institution) and partially from parents financing, while privately-managed preschools are fully financed by parents. In Serbia, two mayor modalities exist in service offer – elective (at the age of 1-5) and mandatory for the age of 6 (this age children have education prepared according to the needs of primary school admission).

Information technologies (IT) at preschools could support many aspects of the institution functional operations, such as educational process and operations management etc. There are many studies on using computers in early childhood education, such as [1], [2], [3], [4], [5]. Other reports are related to using IT in preschool organization management, such as cloud-technology support [6]. Important aspect of IT use in general service-oriented organization management is support to cooperation (informing, interaction) with clients. In the case of preschools children and parents could be considered as clients. Empowering parents in be involved in schools improves efficiency of educational process [7]. It is extremely important to establish and improve communication between parents and teachers in schools [8].

This paper aim is to present results in determination of information needs for the design of preschool web portal, results of existing solutions analysis, as well as to describe the developed web application for the needs of Preschool institution Zrenjanin, Serbia. In conclusion, further improvements plan is described.

2. Related Work

Education-related portals are developed for support to diversity of educational levels (starting from preschool to university), as well as to variety of users (students, teachers etc). They are mainly designed according to national policy for using ICT in education, to support distance learning, knowledge representation and sharing, as well as to support teachers and students network communities [9].

It is of a great importance to define user types and consider their information needs, which influences design criteria and directives, especially for children's web portals [10]. Special attention is given to web-based solutions to support working with children with special needs and for teachers' education [11].

Preschool web portals are developed in many countries. Research [12] presents analysis of Finland municipalities' early childhood education web portals, from the aspect of service quality. There are several quality criteria for analysis:

- Usability (ability to locate needed information without difficulty, clarity, readability, search facilities),
- Usefulness of content (relevant information to user needs, valuable tips on services, customized information presentation, up-to-date information),
- Adequacy of information (accuracy and completeness of information)
- Accessibility (availability and responsiveness, speed of web page loading)
- Interaction (interaction between user and web site in both directions, possibility of setting directions, available guidance, suggestions)

Contemporary technology is directed towards simplification of user interface and availability to large population, therefore mobile devices and applications represent contemporary solutions. Preschool web portal and mobile services development, i.e. mobile phone application development for the Preschool portal needs is represented in [13].

3. Existing Solutions of Preschool Web Portals Analysis

Analysis of existing solutions for preschool web portals IS conducted in general, regarding quality criteria presented in previous section of this paper. Population for analysis is selected among several available solutions in Serbia. Preliminary analysis results will be presented descriptive, considering small sample of available web sites.

Preliminary analysis of existing solutions for preschool portals leads to following quality criteria analysis results:

- ACCESIBILITY - Serbian preschool institutions information are generally available via:
  - Official city or municipality web site includes general static information about preschool institution (Figure 1, [14]).
  - Centralized web portals for the whole preschool institutions, that include information of all particular preschools that are managed and supplied from the central preschool institution (Figure 2, [15])
  - Each preschool has its own web site.
- ADEQUACY of information:
  - Completeness – most centralized Preschool Institutions have similar content that is organized for parents' and suppliers' information needs. Parents-

related info includes contact, organization and news sections, while suppliers-related info includes “Public procurement contests”.

- Accuracy – most information are static, except for news and public procurements sections. Static info is risky, since there is possibility that they not updated regularly and stored/retrieved from database.

- USABILITY – information are clear, readable and it is easy to access them via simple menus or drop-down menus for additional info. Search facilities are usually not available.
- USEFULNESS – information is useful to two target populations of web portal users, but there is no customization or personalization of presented content.
- INTERACTION – there is only one-way interaction, i.e. static presentation of content is available, without any interaction support.

Fig. 1. Official web portal of Preschool institution in Novi Sad announced within the city of Novi Sad portal [14]

Fig. 2. Official web portal of Preschool institution in New Belgrade [15]

4. Development of Preschool Institution Zrenjanin Web Portal

Development of Preschool Institution Zrenjanin Web portal started several years ago as WordPress [16] web application (presented at Figure 3).

Fig. 3. WordPress version of Preschool Institution Zrenjanin Web Portal [17]

Initiation of new web portal development - In December 2014 Preschool Institution Zrenjanin experienced lack of IT staff that could continue WordPress administration and expressed the need for additional features development. Urgent need for publishing data related to "Public procurement contests" lead to the initiation of new website design and implementation within cooperation with University of Novi Sad Technical faculty "Mihajlo Pupin" Zrenjanin.

Analysis of existing solution - The development of new web portal started with the analysis of existing solution functional and non-functional characteristics that were agreed to be supported with new solution, as well as to be enhanced in some parts, if possible. These characteristics are:

- FUNCTIONAL CHARACTERISTICS:
  - Content Management System (CMS) that enable adding or changing data to dynamically supported pages
  - Main sections / menu items: Home, Public procurement, Events, Preschool list, Advices, Image Gallery, About preschool institution, Contact
  - Subsections / submenu items: recent posts, archive
- NON-FUNCTIONAL CHARACTERISTICS:
  - Cyrillic letters
  - Simple horizontal menus and vertical sub-menus (Vertical sub-menus are included within pages at right side).
  - Each post includes title, texts and images and have additional info about post date, keywords of other sections that the post could be categorized, person that posted
  - Background image created by preschool institution art teacher
  - Technology – Wordpress, i.e. php/mysql hosting support

Initial requirements for new solution included development of web application that should support:

- FUNCTIONAL REQUIREMENTS:
  - Dynamic content support – CMS should enable texts, documents and images data entry for each post in any section.
  - All main sections from previous version should be supported in new version.

- Sections that should be supported with dynamic content (saved in database and enabled to be edited) are News pages (2 separate pages – "Priority news" at first page and "Events" at separate page) and "Public procurement contests" page.
- Sections that should be supported with the static content should contain information about central preschool institution and for each preschool (image, location, contact data), image gallery, suggestions/advices to parents.

➢ NON-FUNCTIONAL REQUIREMENTS:

- All content should be in Cyrillic letters.
- Visual design and information (menus) organization similar to previous WordPress web site, with: a) the main menu items arranged in the same or similar order as previous solution, b) use of simple menus (without dropdown submenus) and submenus within pages at right or left side.
- For each post is important to publish date of the post publishing, together with the title, text, images and the attached documents.
- Using the same or transformed background image for all portal pages, respecting the artwork of the preschool institution art teacher presented at previous web portal solution.
- Since existing hosting supports PHP/MySQL technology, future solution should be developed using the hosted and supported technology.

The developed new solution is available at Zrenjanin preschool institution official web site [18] and the public part (the presentational segment) of the web site is presented at Figure 4. Software functions that are supported with new web portal include sections:

1st page- Priority news (with important documents and contests) – menu item "Pocetna"
2nd page - History of institution establishment, organization description – menu item "O nama"
3rd page - Events news (with texts and images about events)- menu item "Desavanja"
4th page - Public procurement contests (with texts and documents) – menu item "Javne nabavke"
5th page - List of all preschools with contact, map location and image – menu item "Vrtici"
6th page - Advices for parents – menu item "Saveti"
7th page - Image gallery – menu item "Galerija"
8th page - Contact information – menu item "Kontakt".

Fig. 4. User interface of new Preschool Institution Zrenjanin Web Portal [18]

After starting option “Prijava” (upper right corner above main menu), web portal administrator could start CMS part with options to enable adding and uploading new content, as well as deleting unnecessary content. All dynamic sections are treated the same way - as news with attachments (images or/and documents) (Figure 5).

Fig. 5. CMS for dynamic part data services at Preschool Institution Zrenjanin Web Portal [18]

5. Stakeholder needs and the solution support analysis

Previously presented new solution was developed as new implementation of existing software functionality. It does not have full functionality of a portal, since it does not support the needs for all stakeholders and the personalized content management for each of the user role. The determination of the complete list of stakeholders and their information needs require a detailed study that should be conducted systematically. Table 1. presents only some most important stakeholders and their needs for the information and functionality that should be supported with preschool web portal, as well the analysis of existing solution support. This table could be basis for further web portal improvements.

Table 1. Main stakeholders for preschool web portal, needs and the existing support analysis

STAKEHOLDER	**NEEDED INFORMATION AND FUNCTIONALITY**	**EXISTING SOLUTION SUPPORT**
Parents	Contests info for preschool admission, other children related contests, events announcements, past events information	Supported
	Specific every-day organization-related info for particular preschool, contact with teachers, reporting issues	Not supported
Preschool teachers	Internal preschool organization info	Not supported
	Professional education of teachers	Not supported
	Contact with parents in organization and issues resolving	Not supported
Preschool management	Getting and sending organization-related data from each preschool to central management institution and vice versa.	Partly supported
Supply companies	Public procurement contests – documents, procedures, decisions	Supported

6. Conclusion

New solution of Preschool Institution Web Portal was developed by using the available hosting technology PHP/MySQL in December 2014 and January 2015. New preschool web portal was initially required to be developed as another implementation of existing solution (with similar data and functionality), that was supposed to solve the problem of content management. New solution, created in raw PHP/MYSQL, replaced previous WordPress solution. All important information from previous solution were organized similarly, but content management is improved. This way, new solution brought new CMS with Serbian user interface and better support to flexibility in the expected changes of the user interface and functionality.

In the period of 2015-2017 the developed CMS was used by deputy director of Preschool Institution Zrenjanin for the data administration. Experiences in using the developed solution lead to the explicitly expressed needs for changes i.e. adding new features that should be developed in new version. These additional requirements are:

- Transition of static-content sections/pages to dynamic, such as image gallery, by adequate support within CMS and public presentation part;
- Adding new section "Document management" for sharing important documents with the portal users.
- Organization of data as monthly arranged archives (supported within previous, WordPress solution).
- Editing feature for all data within CMS (currently only adding and deleting is supported).
- Data export and optimization, because of the limited hosting space.

Additional enhancements, that are also planned, could include:

- Interactive parts for other web portal stakeholders/user profiles. This improvement is made possible with recent infrastructure improvement (by enabling each preschool in Zrenjanin to have internet access and computer equipment).
- Visual redesign, in aim to enable responsive content suitable for mobile devices presentation.

Limitations to further web portal improvements are in current web hosting space and database support, which could be easily solved by selecting another available hosting service. Other limitations for further portal improvement include development team availability, as well as the need for additional engagements in collaboration with administration and teaching staff for requirements gathering and the system testing. The success of the portal development project is very much related to the motivation and knowledge/skills of potential users, which largely depends on their additional computer-usage, internet-usage and the portal-usage training.

7. References

1. Lydia Plowman L, Stephen C: Children, Play and Computers in in Pre-School Education. British Journal of Educational Technology, Vol 36, No 2, 145-157. (2005)
2. Lindahl, M. and Folkesson, A.-M: ICT in preschool: friend or foe? - The significance of norms in a changing practice. International Journal of Early Years Education, Vol. 20, No 4, 422-436. (2012).

3. Lindahl, M. and Folkesson, A.-M: Can we let computers change practice? Educators' interpretations of preschool tradition. Computers in Human Behavior, Vol 28, No 5, 1728-1737 (2012)
4. Chen J-Q, Chang C: Using computers in early childhood classrooms, Journal of Early Childhood Research, Vol 4, No 2 (2006).
5. Vernadakis N, Avgerinos A, Tsitskari E, Zachopoulou E: The Use of Computer Assisted Instruction in Preschool Education: Making Teaching Meaningful. Early Childhood Education Journal. Vol. 33, No 2, 99–104 (2005)
6. Bogdan, V. O. "Optimizing the preschool management by the tools of cloud technologies [Electronic resource]." Reporting conference of the Institute of Information Technologies and learning Tools of NAES of Ukraine. 2015.
7. Delgado-Gaitan C: Involving Parents in the Schools: A Process of Empowerment. American Journal of Education.American Journal of Education. Vol. 100, No. 1, 20-46 (1991)
8. Rimm-Kaufman S.E, Pianta R.C.: Family-School Communication in Preschool and Kindergarten in the Context of a Relationship-Enhancing Intervention, Early Education and Development Journal, Vol. 16, No.3, 287-316, (2010)
9. Sigalov A, Skuratov A: Educational Portals and Open Educational Resources in the Russian Federation, UNESCO Institute for Information Technologies in Education, Moscow, Russia (2012)
10. Large A, Beheshti J, Rahman T: Design Criteria for Children's Web Portals: The Users Speak Out. Journal of the American Society for Information Science and Technology. Vol 53, No 2,79–94. (2002)
11. Kinzie M.B., Whitaker S. D., Neesen K, Kelley M, Matera M, Pianta R.C.: Innovative Web-based Professional Development for Teachers of At-Risk Preschool Children, Journal of Educational Technology and Society, Vol 9, No 4, 194-204 (2006)
12. Koskivaara E, Pihlaja P: Service Quality of Early Childhood Education Web Portals in Finnish Municipalities. In IFIP International Federation for Information Processing, Volume 252, Integration and Innovation Orient to E-Society, Volume 2, eds. Wang W (Boston: Springer). 382-389. (2007)
13. Backvall P: Preschool Evolved – Communication with Preschool Through Mobile Services. Uppsala Universitet. Examensarbete. (2011)
14. Presentation of Preschool institution of Novi Sad within official web site of Novi Sad City, http://www.novisad.rs/predshkolska-ustanova-radosno-deti-stvo [accessed: 29th August 2017]
15. Official web portal of New Belgrade Preschool Institution: http://11april-nbgd.edu.rs/ [accessed: 29th August 2017]
16. Official WordPress website: https://wordpress.com/ [accessed: 29th August 2017]
17. WordPress version of Preschool Institution Zrenjanin Web Portal: http://predskolskazr.edu.rs/indexstaro.php [accessed: 29th August 2017]
18. Official Preschool Institution Zrenjanin Web Portal: http://predskolskazr.edu.rs [accessed: 29th August 2017]

Security Mechanisms for Wireless Multimedia Sensor Networks: A Survey

Angel Dimoski, Zoran Kotevski, Nikola Rendevski

"St. Kliment Ohridski" University – Bitola,
Faculty of information and communication technologies
Partizanska bb, 7000 Bitola, Macedonia
{zoran.kotevski, nikola.rendevski, angel.dimoski}@fikt.edu.mk

Abstract. The technological advances of microelectromechanical (MEMS) systems created novel developmental horizons for powerful sensor-based distributed intelligent systems, capable of retrieving various multimedia content such video and audio streams, still images and scalar sensor data from the physical environment. This wirelessly networked paradigm is called Wireless Multimedia Sensor Networks (WMSNs). WMSNs are designed for both real-time mission-critical applications, which demand strict quality of service (QoS) requirements, such low delay, high throughput and reliability, as well as non-real-time applications which require medium bandwidth and allow certain loss tolerance. In both application scenarios, a concern of great importance nowadays is the data security of WMSNs. As the number of application and usage models rely on the concept of WMSNs, significant research and development attention is dedicated to various mechanisms towards providing privacy and security of such systems. Besides the well-known methods based on cryptography algorithms for data protection, different security challenges are required for each encryption techniques, mostly as a result on constraints of a different nature. At this context, this work presents a survey on the security mechanisms in WMSNs and their performance, complexity and suitability for implementation in WMSNs.

Keywords: wireless multimedia sensor networks, security, QoS, multimedia encryption.

1. Introduction

WMSNs are networks built from wirelessly connected smart devices capable of capturing video and audio streams, still images, and scalar sensor data in real-time and non-real-time application scenarios. WMSNs development coexist with the recent developmental trends of the next-generation network paradigms, such Internet of Things (IoT), Cyber Physical Systems (CPS), Industry 4.0 and Tactile Internet (TI) as 5G-enabled communication technologies, changing the way of how information may be

generated and transmitted. In such developmental roadmap, the massive production of multi-sensor networked devices, capable to produce and distribute larger and high quality multimedia content is highly expected. Sensor-based systems could span a wide range of application areas, including scientific research, military, disaster recovery and rescue, healthcare, industry and robotics, environmental monitoring, smart homes etc. [1]. In [2], the WMSNs applications are classified in five categories: Surveillance, Traffic Monitoring, Personal and Health Care, Gaming and Environment and Industry. Those applications transmit different types of multimedia data. In some cases the data is transmitted during short periods of time (images), but others require transmission of large amounts of multimedia content in real time (streaming). Considering the wireless transmission in well-known and crowded frequency bands, the risk of unauthorized eavesdropping and changing the data during transmission, open real security challenges for WMSNs [3]. Furthermore, Denial of Service (DoS) attacks are particularly devastating for resource-constrained wireless sensor nodes and also demands appropriate security measures. Among the available security protection mechanisms for WMSN, the conventional cryptography techniques are effective and can provide confidentiality, integrity and authenticity to sensed data and sensor nodes. Cryptography is a science area which provides techniques for protecting data with transformation of original unprotected data into unreadable data structure, which can only be read by an appropriate recipient. Cryptography includes two processes, encryption and decryption, where the transmission of the data decryption key is ciphered as well. Different multimedia services have different QoS and encryption demands [4]. Hence, this paper provides a survey and analysis of the security challenges in the design of WMSN platforms and protocols. In Section 2, we present the existing security challenges in WMSNs context. Cryptography algorithms are surveyed and discussed in Section 3, while the encryption mechanisms for multimedia streaming are presented in section 4. Section 5 concludes the paper with summary of the contribution.

2. Security in Wireless Multimedia Sensor Networks

As the radio waves are considered as transmission media in WMSN, it is well known that such systems are significantly more vulnerable to attacks than the wired networks [3]. The increased use of sensing devices with communication capabilities and limited processing power, produces security threats that can harm the applications in many ways, as the resource-constrained nature of sensor nodes plays a central role in the implementation of security mechanisms. According to [6], where the security requirements are discussed, and most WMSNs applications are affected by at least one of the following: authenticity, availability, confidentiality, data freshness, integrity and localization. Authentication mechanisms, secure localization algorithms, trust management, privacy aspects, secure compression and aggregation algorithms are presented in [8]. Besides the existing security mechanisms and well known technologies widely implemented in other networking technologies, one question naturally rises: what is the exact justification of security analyses of WMSNs as a result of its characteristic architecture and types of attacks specific for this network paradigm? There are several types of attacks in WMSN, which can compromise security requirements. Some of the attacks can be avoided or minimized, but some of them may

not be easily avoided such as intrusion attacks and DoS attacks. Attackers are looking for vulnerabilities, which could be found in some communication layers, and they can perform eavesdropping on the transmission, altering confidential data or prejudicing the network operation with the insertion of malicious information. Security attacks in WMSN may be of four different types: interruption, interception, modification and fabrication. Differences between them are: interruption attacks compromise availability; interception compromise confidentiality; modification prejudices integrity; and fabrication impairs authentication [9]. In such a way, security defenses are often required, and there are many ways to protect wireless multimedia sensor networks.

In a wireless sensor network which supports multimedia streaming applications, denial of service (DoS) and service disruptions can become severe problems. An adversary can corrupt or inject false multimedia packets, and these packets may be forwarded all the way to the base station where they are found to be unusable. This sensor nodes would be drained much faster of battery energy [7]. It is expected that DoS will be the main attack to be worried about in most of the WMSN. DoS attacks represent quite complex problems, since they can be performed in many different ways, and against any of the different communication layers. In [10] different DoS attacks are presented (Physical attacks, Tampering attacks, Jamming attacks, Collision attacks, MAC protocol attacks, Routing Protocol attacks, Transport layer protocols attacks and Intrusion attacks. With Eavesdropping attacks the wireless communication channel and the data can be discovered and used to attack the privacy of individuals referred to by those, by sniffing the messages exchanged by the network nodes. Masquerading provides data retrieving by means of some malicious nodes that misroute the packets and mask their real nature behind the identity of nodes that are authorized to take part in the communications. For eavesdropping and masquerading attacks there are three different types of solutions: anonymity mechanisms based on data cloaking, privacy aware mechanisms based on secure communication channels and privacy policy based approaches [8].

Security defenses may be focused on the network or on the data. When protecting the network, secure protocols may be used to avoid attacks, as denial of service, man-in-the-middle and general packet redirection. But, concerning the data the most effective approach is cryptography, which is the basic defense mechanism in wireless sensor networks that directly protects the data. Cryptography is a set of techniques for transforming original information into a set of unreadable data, allowing it to be read only by the correct recipients [6, 11, 12]. Because of the constrains that are inherent to wireless sensor networks, especially the ones with high computing power and communication overhead, traditional cryptography may not be feasible for WMSN, and therefore optimized cryptography is employed. Cryptography provides authenticity, confidentiality and integrity for wireless multimedia sensor networks. The use of cryptography keys provides authentication of source nodes, where such keys would be required to recover the original data and confidentiality assured as well. Furthermore, if the original information cannot be accessed, it cannot be adulterated, which adds to the provision of integrity. Authors in [7] concluded that all existing multimedia encryption schemes are based on three mechanisms: position permutation, value transformation and combination. In [5], security is divided in four categories: Efficient management of Quality of Experience (QoE) and Quality of Service (QoS), Privacy, Authentication and Node Localization.

3. Cryptography Algorithms

There are numerous cryptography algorithms available for various purposes. Some of those algorithms are directly used or adapted. Popular cryptography algorithms for WMSN include:

The symmetric encryption provides a single shared key for both encryption and decryption functions. As a result, the process of cryptography is easier to implement. However, the biggest challenge is how to securely distribute the shared key (AES, DES, IDEA). The standard symmetric encryption schemes, such as DES and AES, are commonly used. However, these schemes are unsuitable for multimedia data. Multimedia data is generally larger in size and use of these symmetric encryption schemes has memory and computation requirements not suitable for the sensor nodes, because its real time nature requires faster encryption.

The asymmetric encryption uses a pair of key stoppers to perform data encryption. A public key that is known by all nodes of the network is used to encrypt data, and a private key known only by the destination node is used to decrypt that data (RSA, ECC). Authors in many papers noted that asymmetric cryptography is not appropriate for WSN, because of its needs for more resource power and more processing time. Other papers have proven inversely. In the key based security management, if the decryption key is lost or corrupted during transmission, it is difficult to recover the information [13]. WMSNs asymmetric cryptography will be used more than symmetric cryptography because it makes a lot easier to solve several security problems related to eavesdropping and compromised nodes [3].

4. Multimedia Streaming and encryption

Streaming media is video or audio content sent in compressed form over the network and displayed immediately, rather than being saved to the secondary memories. Unlike scalar sensor networks, multimedia data include snapshots or streaming multimedia content. Snapshot multimedia data contains pictures obtained in a short period of time (still images). Streaming multimedia content is obtained over longer time periods and needs to be delivered in real time [2]. Video streaming supports military operations by delivering critical information rapidly and dependably to the right individual or organization at the right time. This improves the efficiency of combat operations. The new technologies must be integrated quickly to meet the requirements of present time. The same is for traffic monitoring [1]. In [14] the authors present problems in multimedia streaming where they elaborate that major problems when WMSN is used in surveillance are the lack of storage to save and record data, and the lack of services and battery consumption. Because of the lack of services the problems occur in the transfer of data, like data loss, quality degrade and low transfer rates. The authors also present some solutions to the aforementioned problems, and emphasize that the biggest problem is how to secure the data during transmission.

4.1. Multimedia Encryption Techniques

Multimedia sensors can sense and transmit data in the form of image, video and audio. As more data has to be encrypted and decrypted, cryptography naturally becomes more complex and resource-demanding in WMSNs. Because of the complex compression operations, the distributed environment and the limited bandwidth and power resources of WMSNs, there are inevitable needs for aggregation algorithms capable to decrease the total amount of information to transmit [15]. Images, video and audio have different particularities, and naturally, the cryptography mechanisms are different. In general, secure data transmissions can be achieved through symmetric or asymmetric cryptography, which may be performed through different algorithms. Both of those cryptography algorithms have pros and cons. Moreover, in the following text, multimedia cryptography techniques are presented, for each media type separately, as well as the most commonly used techniques.

4.2. Image Cryptography Techniques

Image sensors generate large amount of data traffic in a WMSNs, which may load the network bandwidth. The image data transmissions in WSNs can significantly degrade the network performance and sensor lifetime. The concept of secret sharing is used to develop a secure routing protocol. It works by dividing data packets (image) into smaller packets called shares, and these shares are sent through disjoint multipaths. An unauthorized user has to intercept at least a threshold number of those shares before the packet can be decrypted. In such case, unauthorized users must compromise all paths in order to decrypt the message. However, the drawback of this scheme is that the disjoint paths have to be determined before the shares are transmitted [5, 20]. Sending shares over already determined paths may not guarantee a video quality. In [16] authors are presenting data fusion techniques with algorithm of compressive sensing and watermarking, grouped in three data fusion categories: low-level fusion, medium-level fusion and high-level fusion.

4.3. Video Cryptography Techniques

Key challenges for video coding in the sensor nodes are also the low power and computational capabilities. For this reason, video sensors typically employ compression techniques based on coding mechanisms [7]. The videos are demanding more requirements of processing, memory, energy consumption, transmission delay and jitter, and more bandwidth than image transmissions. Their quality will depend on the resolution, the frame rate, the color pattern and the similarity between the reconstructed and original (source) video. These variables add complexity to the cryptography. An encryption algorithm for video streaming need to have at least two characteristics: the encryption time should be low to avoid delays and the compression rate of the video should not be decreased [21].

The idea of joint coding is to integrate encryption into compression operation by parameterization of the compression blocks, without modifying the compressed bits. Two main compression blocks, where these techniques have been applied, are Wavelet Transform and Entropy Coding. Advantages of JVCE is that, it compresses and encrypts information in a single operation, making it feasible for mobile and embedded devices to ensure multimedia security with their low power consumption. Considering this fact, JVCE reduce the latency of encryption operation which is useful for real-time video delivery. The fastest algorithm, like AES, is computationally very intensive for many of the real-time multimedia data.

Video scrambling method uses filter banks or frequency converters, and it is performing permutation of the signal in time domain or distortion of the signal in the frequency domain. This algorithm offers less security, and this method can be easily compromised. Other video coding algorithms may also be used for optimized cryptography [22]. The Distributed Video Coding (DVC) uses a video compression technique with low encoder complexity. The encoder can be very simple, but the decoder is significantly more complex [23].

Security and video quality are progressively significant attributes for wireless multimedia sensor networks. However, it is crucial to integrate security and video quality together for video transmission as delivering video data across a secure path does not often meet the video quality requirements in many traditional approaches.

4.4. Audio Cryptography Techniques

The way of how cryptography is performed depends on audio compression requirements, and in general, audio data compression requires less processing compared to the compression of images and video streams.

Pulse code modulation (PCM) is a technique that can be quite effectively encrypted, as both encryption and decryption are much simpler to execute. In this method, the audio signal samples will be encrypted using very common arithmetic algorithms and the total number of bits will be kept unaffected. The bandwidth required to transmit the encrypted signal is same as for the original signal. In the receiver section, the encrypted signal is decrypted using exactly the reverse algorithm used earlier for encryption [24].

CVSD is a type of delta modulation in which the step size of the approximated signal is progressively increased or decreased, as required, to make the approximated signal close match with the input analog wave.

Modified Discrete Cosine Transform (MDCT) is designed to be performed on consecutive blocks of a larger dataset, where subsequent blocks are overlapped so that the last half of one block coincides with the first half of the next block. This overlapping, in addition to the energy-compaction qualities of the DCT, makes the MDCT especially attractive for signal compression applications, since it helps to avoid artifacts stemming from the block boundaries [6, 25]. MDCT is employed in most modern lossy audio formats, including MP3, AC-3, Windows Media Audio, AAC. Similarly as with images and videos, audio streaming in wireless sensor networks can also be protected with watermarking.

4.5. Common cryptography techniques for all media types

Encrypting large size multimedia data might suffer high computation complexity and latency. The secret information cannot be recovered if the decryption key is lost or the encrypted content is corrupted during transmission [13].

The encryption of multimedia data may be very costly in time and computing power, which may be infeasible for some sensor networks [16]. To avoid this constrains, using selective encryption is the most appropriate way. In selective encryption, the basic idea is to encode only a set of blocks of sensed images. There are two coding algorithms which are well suited for selective encryption: quad-tree coding and wavelet coding [18]. Quad-tree coding is based on a computational rooted tree [4], which separate the original image into different sub-quadrants, which means that every node has zero or four sub-quadrants and reconstruction process is related to their position in the tree. While at DWT-based (Discrete Wavelet Transform) algorithms creates a hierarchy of frequency bands. DWT decomposes images into smaller parts, called sub-bands, each sub-band have different importance in original image reconstruction process [19]. Selective encryption technique is combining compression with encryption. This technique can handle real time audio and video data efficiently. This method selects only the most important coefficients of a compression process and encrypts those coefficients. Coefficients which are less important are not encrypted [17, 22]. To optimize the cryptography of video streams in wireless sensor networks selective encryption is used as well.

Another security mechanism used in protecting multimedia data is watermarking which provides authentication. Watermarking process hides authentication information in the original data. A digital watermark is a special marker that is embedded into scalar, audio, image or video data, aiming to provide a mechanism to identify ownership and copyright. The objective of digital watermarking is to protect the intellectual property of multimedia contents such as copyright protection, contents archiving, metadata insertion, broadcast monitoring, tamper detection and digital fingerprinting [15]. Watermarking system has two components: embedder (3 inputs: cover data, watermark, and key) and detector (1 input: key). In general, any media transmission over wireless sensor networks may be protected using watermarks. Also, watermarking is used in combination with others security mechanisms. However, watermarking solutions might be vulnerable to attacks from entities that know how the watermarks are done [3].

5. Conclusion

In this paper, we have surveyed the security threats and defending mechanisms in Wireless Multimedia Sensor Networks. WMSNs have been deployed in military sensing, traffic surveillance, target tracking, monitoring, and healthcare. We argue that DoS attacks are challenging problem in WMSN security. In this context, cryptography will play a central role in WMSN. There is symmetric and asymmetric encryption, and there are different advantages and drawbacks between them. Asymmetric cryptography will be used more than symmetric cryptography because it makes a lot easier to solve

several security problems related to eavesdropping and compromised nodes. Images, videos and audios have different particularities, and cryptography is different for each media. The common encryption techniques like selective cryptography and watermarking have advantageous, but serious drawbacks for implementation in WMSNs exist. Considering the analyses conducted and the literature review, selective cryptography technique appears as the most appropriate approach suitable for securing data in WMSNs.

6. References

1. Vivek Katiyar, Narottam Chand, Naveen Chauhan;Recent advances and future trends in Wireless Sensor Networks,INTERNATIONAL JOURNAL OF APPLIED ENGINEERING RESEARCH, DINDIGUL Volume 1, Number (2010)
2. Ian F. Akyildiz, Tommaso Melodia, Kaushik R. Chowdhury; Wireless Multimedia Sensor Networks: Applications and Testbeds, Proceedings of the IEEE (2008)
3. Almalkawi, I.; Zapata, M.; Al-Karaki, J.; Morillo-Pozo, J. Wireless multimedia sensor networks: Current trends and future directions. Sensors (2010)
4. De Oliveira Gonçalves, D.; Costa, D.G. A Survey of Image Security in Wireless Sensor Networks. J. Imaging (2015)
5. Harjito, B.; Han, S. Wireless Multimedia Sensor Networks Applications and Security Challenges. In Proceedings of the International Conference on Broadband, Wireless Computing, Communication and Applications, Fukuoka, Japan (2010)
6. Daniel G. Costa *, Solenir Figuerêdo and Gledson Oliveira; Cryptography in Wireless Multimedia Sensor Networks: A Survey and Research Directions; MDPI (2017)
7. Satyajayant Misra, Martin Reisslein, and Guoliang Xue; A Survey of Multimedia Streaming in Wireless Sensor Networks, IEEE Communications Surveya & Tutorials, Volume 10, Number 4 (2008)
8. Luigi Alfredo Grieco, Gennaro Boggia, Sabrina Sicari, Pietro Colombo; Secure WMSN, Third International Conference on Mobile Ubiquitous Computing Systems, Services and Technologies (2009)
9. Wang, Y.; Attebury, G.; Ramamurthy, B. Security issues in wireless sensor networks: A survey. Int. J. Future Gener. Commun. Netw (2013)
10. Guerrero-Zapata·Ruken Zilan· José M. Barceló-Ordinas·Kemal Bicakci·Bulent Tavli; The future of security in Wireless Multimedia Sensor Networks; Published online: 3 December 2009 © Springer Science+Business Media (2009)
11. Sen, J. A Survey on Wireless Sensor Network Security. International Journal on Commun. Netw. Inf. Secur. (2009)
12. Modares, H.; Salleh, R.; Moravejosharieh, A. Overview of security issues in wireless sensor networks. In Proceedings of the International Conference on Computational Intelligence, Modelling & Simulation, Langkawi, Malaysia, (2011)
13. Akshaya Gayathri, Ms. Suganthi, Saranya; Energy Efficient Image Transmission with Security in Wireless Sensor Networks, International Journal of Computer Applications in Engineering Sciences, Vol 3, Number 1(2013)
14. Himanshu Diwan, Pooja Agrawal, A.K.Dwivedi, Current Status and Design Challenges in Wireless Multimedia Sensor Networks, International Journal of Engineering Trends and Technology (IJETT) – Volume 6, Number 2 (2013)
15. Bambang Harjito, Elizabeth Chang; Secure communication in WMSN using watermarking; IEEE International Conference on Digital Ecosystems and Technologies (DEST), (2010)

16. Rui Gao, Yingyou Wen, Hong Zao; Secure Data Fusion in WMSN via Compressed Sensing, Journal of Sensors (2015)
17. Viral Patel, Krunal Panchal; Survey on Security in Multimedia Traffic in Wireless Sensor Network; IJEDR Volume 2, Issue 4, ISSN: 2321-9939 (2014)
18. Bhimrao S Patil Dept of CSE BKIT, Bhalki, Karnataka, INDIA; Image Security in Wireless Sensor Networks using wavelet coding; International Journal on Emerging Technologies - Special Issue on NCRIET (2015)
19. Costa,D.G.; Guedes,L.A. A discrete wavelet transform (DWT)- based energy-efficient selective retransmission mechanism for wireless image sensor networks. J.Sens Actuator Netw. (2012)
20. Rashwan, Honggang Wang, Dalei Wu, Xinming Huang; Security–quality aware routing for wireless multimedia sensor networks using secret sharing, SECURITY AND COMMUNICATION NETWORKS, (2015)
21. Varalakshmi, L.M.; Sudha, G.F.; Jaikishan, G. A selective encryption and energy efficient clustering scheme for video streaming in wireless sensor networks. Telecommun. Syst. (2014).
22. Pande A, Zambreno J. Embedded Multimedia Security Systems Algorithms and Architectures, (2013)
23. Bernd Girod (Fellow IEEE), Anne Margot Aaron, Shantanu Rane and David Rebollo-Monedero; Disturbed Video Coding, Proceedings of the IEEE 93 (2005)
24. Avijit Hira, Nazmus Sakib, Nayan Sarker; PCM based audio signal security system; International Conference on Advances in Electrical Engineering (ICAEE), (2013)
25. Honggang Wang, Michael Hempel, Dongming Peng, Wei Wang, Hamid Sharif, Hsiao-Hwa Chen; Index Based Selective Encrypton for WMSN; IEEE Transactions on Multimedia April (2010)

Model for implementing Big Data in Smart City environments

Nenad Kermeci[1], Dalibor Dobrilovic[2], and Milan Malic[3]

[1] Toronto Dominion Bank, Data Strategy,
79 Wellington West, Toronto, Ontario, Canada, M5K 1A1
nkermeci@rogers.com
[2] University of Novi Sad / Technical Faculty "Mihajlo Pupin", Djure Djakovica bb,
23000 Zrenjanin, Serbia
dalibor.dobrilovic@uns.ac.rs
[3] Panonit, 21000 Novi Sad, 7 Miroslava Antića, Republic of Serbia
milanmalic@outlook.com

Abstract. Smart City technologies have become reality in recent years. More and more cities all over the world start to implement Smart City technology with variety of applications each year. Considering the large quantity of data that can be collected in Smart City systems, the implementation of Big Data in those environments becomes more realistic solution. This paper is presenting Big Data implementation model in the Smart City environment. Research is focused on possible models and technologies for implementing Big Data in Smart City environments.

Keywords: Smart City, Big Data, Internet of Things, Hadoop.

1. Introduction

In recent years Smart City technologies become reality in cities all over the world. A large number of cities improve quality of life for their citizens, and enhance management by implementation of emerging smart technologies. In those environments, a variety of large quantity sensor acquired data can be collected in Smart City systems, and used for the implementation of Big Data technology.

This paper is presenting Big Data implementation model in the Smart City environment. Research is focused on possible models and technologies for implementing Big Data in Smart City environments. The first phase of the research is based on exploring applicability of emerging wireless sub-Gigahertz technologies such as LoRa and LoRaWAN. In the second phase, data storage, further data analyses and processing of acquired data is explored.

This paper is structured as follows: after the introduction, the short part about Smart City environment and Low-Power Wide-Area Network (LPWAN) technologies basics (LoRa/LoRAWAN) is presented. The model for sensor deployment and data collection

is given in the next section. The following subsection describes analyses of four tier architecture for Big Data model implementation in Smart Cities, and the last subsection gives the implementation of Hadoop Distributed File System. Prior the conclusion and further work section, the Hadoop ecosystem is briefly described.

2. Smart City environments

In the beginning of the paper it is important to explain the term "smart city". This term defines the new urban environment, especially designed for performance through information and communication technologies (ICTs) and other forms of physical capital. With the effective management of resources through intelligent management, visionaries hope that cities will drive a higher quality of life for their citizens, drive down waste, and improve economic conditions" [1]. According to [2, 3] the relevant goals for a smart city are: Smart mobility (traffic management, bike/car/van sharing, multimodal transport, road conditioning monitoring, parking system, route planning, electric car gearing services); Smart grid/energy (power generation / distribution / storage, energy management, smart metering, street lightening optimization); Public safety (video / radar / satellite surveillance, environmental and territorial monitoring, children protection); Smart governance (transparent decisional process, greater involvement of citizens in legislative initiatives, public-private partnerships, online taxing systems); Smart economy (high-level jobs, competitiveness, entrepreneurial spirit, innovation and research in the field) and Smart life (cultural and educational facilities, meaningful events, entertainment and guided tours, access to cultural sights and historical monuments, good conditions for health).

3. Low-Power Wide-Area Network (LPWAN) technologies

Low-Power Wide-Area Network (LPWAN) is wireless telecommunication network designed for long range communications at a low bit rate [4]. The technology is designed primarily for battery operated sensor stations. LPWAN represents very suitable Smart City technology, enabling wireless connectivity in urban areas. LoRa, LoRaWAN™ and Sigfox are LPWAN technologies. The ongoing research presented in this paper in the sensing layer used 868MHz LoRa [3, 5] and LoRaWAN modules for small city scenarios.

There are number of differences in LoRa and LoRaWAN. LoRa contains only physical and link layer protocol and is perfect to be used in P2P communications between nodes, also LoRa modules are a little cheaper. LoRaWAN includes the network layer and its network architecture is star-of-stars topology in which gateway is

a transparent bridge relaying messages between end-devices and a central network server in the backend. Gateways are connected to the network server via standard IP connections while end-devices use single-hop wireless communication to one or many gateways [3, 4].

4. Model of Implementing Big Data in Smart Cities

The model of Big Data implementation in Smart Cities presented in this section is based on [6] and [7]. This model is presented in details, from sensor deployment structure, with four tier Big Data model architecture and finally with the architecture of distributed file system supported with Hadoop.

4.1. Sensor Deployment

The sensor deployment model is based on [6] and it is presented in Figure 1. The sensor deployment model consists of several components. Environmental pollution monitoring system (1) is the system for monitoring sulfur dioxide level, carbon monoxide, ozone and other gases important for measuring air pollution, including particular metals and noise detection as well. Vehicular traffic monitoring system (2) is one of the most significant services for Smart City, because it can help monitoring traffic jams, traffic accidents and improve the citizen health in monitoring vehicular caused gas emissions.

Weather & Water monitoring system (3) provides the meteorological related data like temperature, rain, humidity, barometric pressure, wind speed, UV radiation, as well as water levels in rivers, lakes, dams, and other reservoirs. This system should help in preventing floods during heavy rains, snow melting and dam breakage. Smart Parking system (4) helps in detecting the need for new parking spaces in the specific areas and to offer citizens information on the nearest parking opening in the real-time.

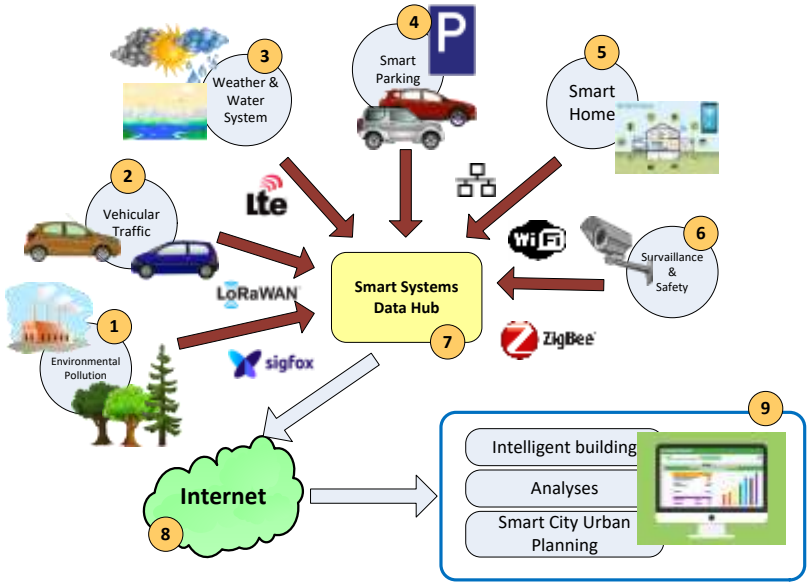


Fig. 1. The sensor deployment model: (1) Environmental pollution monitoring system, (2) Vehicular traffic monitoring system, (3) Weather & Water monitoring system, (4) Smart Parking system, (5) Smart Home @ Smart Building system, (6) Surveillance & Safety system, (7) Smart Systems Data Hub. (8) Internet based infrastructure for data transfer, (9) System for storing, analyzing and planning.

Smart Home @ Smart Building system (5) are used for monitoring in house temperatures, heating management, gas levels, electricity consumption, and smoke and fire detection. At the end Surveillance & Safety system (6) increases security and safety of citizens by uniting security monitoring systems and using it in prevention of robberies, kidnappings, car theft, terrorist attacks and quick police notification and deployment. All enlisted smart systems can be connected with one of the following technologies presented at figure: wireless like SigFox, LoRa, LoRaWAN, WiMAX, ZigBee, IEEE 802.15.4, Wi-Fi, 3G/LTE or wired like Ethernet, etc.

Smart Systems Data Hub (7) in this model is to collect and transfer all sensor data to centrally located smart system. Internet based infrastructure for data transfer (8) has role of fast, reliable and secure data transfer from hub to the System for storing, analyzing and planning (9).

4.2. Four tier Big Data model architecture

Figure 2 is presenting a four tier Big Data model architecture located physically and logically in the centre of the system. The model is based on [7]. Bottom tier has a purpose of data generation and collection. Collection of the data is made by sensor stations, deployed in various smart systems as shown in Figure 1. The collected data is transferred with various communication technologies such as: LoRa, LoRaWAN, WiFi, etc as presented in deployment model in Figure 1. This is the role of Intermediate Tier I.

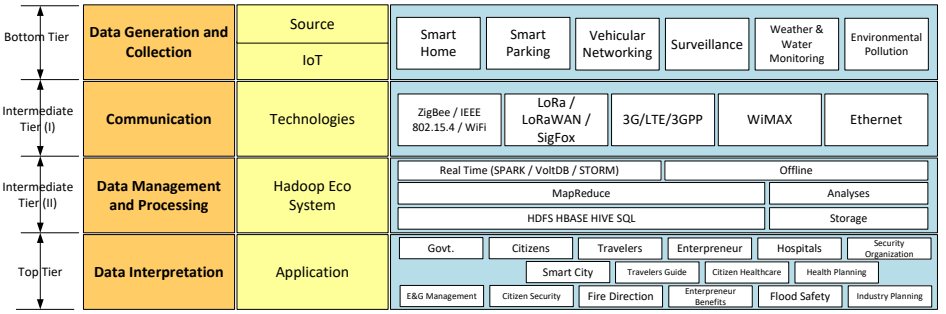


Fig. 2. Four Tier architecture of Big Data model for Smart City systems. Bottom tier is designed for data generation and collection, Intermediate tier I is designed for communication and data transfer technologies, Intermediate tier II is designed for data management, storing, processing and analyses. The last Top Tier is designed for data interpretation and visualization in user applications.

The role of Intermediate Tier II covers the data management, storage, access and processing. All these features are covered within Hadoop Ecosystem for real-time (with the usage of SPARK, VoltDB or STORM) or offline processing. The tier features are additionally enabled at lower layer of Hadoop with the usage of MapReduce. For managing databases: HDFS, HBASE, HIVE or IMPALA can be used. The Top tier has a role of data interpretation and visualization. This tier creates reports based on performed data analyses and makes data available to users through several applications depending of the system (e.g. Smart Home, Smart Parking, etc.).

4.3. Hadoop Ecosystem

Two major components of Hadoop ecosystem, crucial for the presented Smart City - Big Data model are HDFS and MapReduce.

Hadoop Distributed File System (HDFS) is data service for managing files in Big Data environments. Data files are broken into blocks with the sizes between 64-128 MB, and distributed across the nodes. Each block is replicated across multiple, typically three nodes. In HDFS cluster, the nodes can have two functions. Name Node function is responsible of managing the distribution of the blocks. Data Node function is responsible of processing the data blocks. Figure 3 illustrates the HDFS structure. [8]

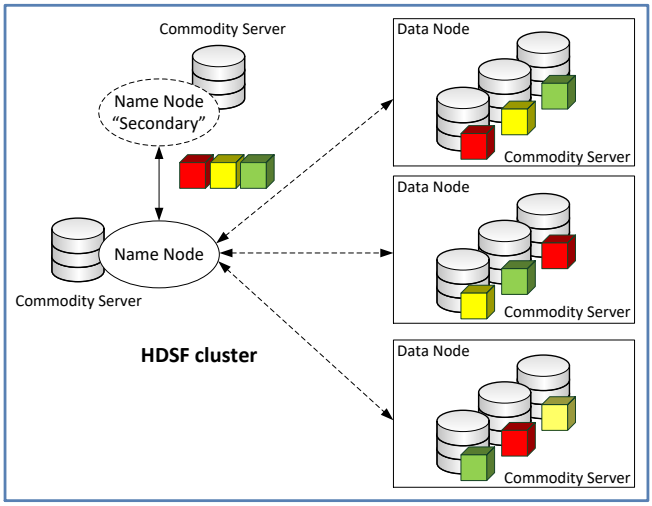


Fig. 3. The structure of the HDFS (Hadoop Distributed File System) with one Name Node and three Data Nodes

Hadoop MapReduce is an engine that takes large quantities of input data and processes it efficiently to produce the output. It operates in two main phases. The first phase is the Map phase. In this phase the input data blocks are received and processed to produce an output in the form of a sorted list of key-and-value pairs. The second phase, is the Reduce phase. In this phase the results from mappers are received, analyzed and merged to produce the final output. In MapReduce, two trackers are used; a JobTracker (for job submission) and TaskTracker (for Map and Reduce process execution). [8]

In this model Hadoop is chosen as a basic building block of the model although some experiments like in [9] revealed that MySQL database outperform Hadoop. Indications are that many third-party applications will be developed on the market for Hadoop data development in the near future, which will greatly improve speed and usability of Hadoop. One example is ETL (Extraction, Transformation and Loading) tool Talend. Hadoop is the largest open-source framework adopted for managing Big Data with Map/Reduce approach [10]. One of the main benefits of Hadoop is its ability

to work in a distributed environment. Due to the nature and structure of Smart City environments, specific gathering process which combines data from multiple locations, the geographical distribution of the databases can be an advantage. Additionally Hadoop has the ability to replicate data for security and reliability.

5. Conclusion and further work

This paper presents the proposal of Big Data model for implementation in Smart City environments. Basics of the Smart City environments are described, along with the communication technologies suitable for the usage in those environments. After the introductory section, implementation Big Data model in the Smart City environment is presented in three levels. First level is sensor deployment model, second level is four tier Big Data model architecture, and third level is HDFS (Hadoop Distributed File System).

This specific research is part of an ongoing bigger project. In this research the experiments and evaluation of applicability of LoRa and LoRaWAN technologies, in small scale Smart City environments are tested. In the second phase the theoretical research for Big Data model implementation in those environments is performed. Based on the research of related works and up to date achievements the model for implementation of Big Data in Smart City system is developed.

In the further work the research will be pointed in two directions. One direction will be focused on creation of sensor stations prototype using LoRa/LoRaWAN data collection communication technologies, and transfer to the central Smart City system. The prototype sensor stations will be developed using the author's previous experiences [11, 12, 13, 14]. The second research direction will be pointed towards the prototype of presented Big Data model for processing collected and stored data.

6. References

1. Stimmel, C. L.: Building Smart Cities Analytics, ICT, and Design Thinking, CRC Press. Boca Raton, FL, USA: Taylor & Francis Group. (2016)
2. Georgescu, M., Popescul, D.: The Importance of Internet of Things Security for Smart Cities Smart Cities Technologies, ExLi4EvA (pp. 3-18): InTech. (2016)
3. Dobrilović, D., Malić, M., Malić, D., Sladojević, S.: Analyses and Optimization of Lee Propagation Model for LorRa 868 MHz Network Deployments in Urban Areas, Journal of

Engineering Management and Competitiveness (JEMC), Vol. 7, No. 1, 2017, 55-62, ISSN 2217-8147 (2017)

4. LoRa Alliance™ Technology, from https://www.lora-alliance.org/What-Is-LoRa/Technology, (2017)
5. Libelium.: Waspmote-LoRa-868MHz_915MHz-SX1272 Networking Guide. from Libelium Comunicaciones Distribuidas S.L. http://www.libelium.com/downloads/documentation/waspmote_lora_868mhz_915mhz_sx1272_networking_guide.pdf (2017)
6. Rathore, M. M., Ahmad, A., Paul, A.: IoT-based smart city development using big data analytical approach, In Proceedings of 2016 IEEE International Conference on Automatica (ICA-ACCA), Curico, 2016, 1-8. doi: 10.1109/ICA-ACCA.2016.7778510 (2016)
7. Rathore, M. M., Ahmad, A., Paul, A., Rho, S.: Urban planning and building smart cities based on the Internet of Things using Big Data analytics, Computer Networks, Volume 101, 2016, 63-80, ISSN 1389-1286, http://dx.doi.org/10.1016/ j.comnet.2015.12.023. (2016)
8. Alshawish, R. A., Alfagih, S. A. M., Musbah, M. S.: Big data applications in smart cities. In Proceedings of the International Conference on Engineering & MIS (ICEMIS), Agadir, 2016, pp. 1-7., doi: 10.1109/ICEMIS.2016.7745338 (2016)
9. Jach, T., Magiera, E., Froelich, W.: Application of HADOOP to Store and Process Big Data Gathered from an Urban Water Distribution System, Procedia Engineering, Volume 119, 2015, 1375-1380, ISSN 1877-7058, http://dx.doi.org/10.1016/ j.proeng.2015.08.988. (2015)
10. Fazio, M., Celesti, A., Puliafito, A., Villari, M.: Big Data Storage in the Cloud for Smart Environment Monitoring, Procedia Computer Science, Volume 52, 2015, 500-506, ISSN 1877-0509, http://dx.doi.org/10.1016/j.procs.2015.05.023. (2015)
11. D. Dobrilovic, B. Odadzic, Z. Stojanov and V. Sinik, "Testing Zigbee RF module applicability for usage in temperature monitoring systems," 2014 22nd Telecommunications Forum Telfor (TELFOR), Belgrade, pp. 415-418. doi: 10.1109/TELFOR.2014.7034436 (2014)
12. D. Dobrilovic, Z. Stojanov and B. Odadzic, "Teaching application development for RFID/ZigBee networks using open source hardware," 2014 X International Symposium on Telecommunications (BIHTEL), Sarajevo, pp. 1-6. doi: 10.1109/BIHTEL.2014.6987641 (2014)
13. D. Dobrilovic, Z. Stojanov, V. Brtka, Z. Čović and N. Bilinac, "Software application for analyzing ZigBee network performance in university courses," 2014 IEEE 12th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, pp. 73-77. doi: 10.1109/SISY.2014.6923560 (2014)
14. D. Dobrilovic, Z. Stojanov, B. Odadzic and V. Sinik, "Platform for teaching communication systems based on open-source hardware," 2015 IEEE Global Engineering Education Conference (EDUCON), Tallinn, pp. 737-741. doi: 10.1109/EDUCON.2015.7096051 (2015)

Pocket-Sized Devices (PSUD) and Smartphones as Ultrasound on Point-Of-Care (POC)

Zoran Ćirić[1], Sara Gračić[2]

[1] Faculty of Economics, Segedinski put 9-11,
24000 Subotica, Serbia
zoran.ciric@ef.uns.ac.rs
[2] PhD student, Faculty of Economics, Segedinski put 9-11,
24000 Subotica, Serbia
saritta4u@gmail.com

Abstract. Medical diagnostics has significantly improved, although some patients are discharged without correct diagnosis, thus increasing the probability of fatal outcome. When expecting patients, doctors rely on paramedic's description which can cause misunderstanding and administering inadequate therapy. Wireless communication, smartphones, pocket-sized ultrasound devices, key chains and other accomplishments of telemedicine are gaining importance abroad. Wireless communication enables remote patient's chart access, examination in real time for adequate preparation of hospital's team. Pocket-sized ultrasound can be important in diagnosing emergency patients, giving proper treatment, thus decreasing the risk of fatality. Medical apps for smartphones are intended for doctors, med student and patients. It seems like these innovations are not given enough attention in Serbia. Through clinical trial results and different experiences of medicals and patients, dilemmas in focus of this research are investigated. The main goal is to modernize existing medical information systems in Serbia, due to economic, social and medical benefits for patients, doctors and government.

Keywords: medical information systems (MedIS), pocket-size ultrasound devices (PSUD), telemedicine, smartphones

1. Introduction

Wireless communication, smartphones, pocket-sized ultrasound devices, key chains and other accomplishments of telemedicine are gaining importance abroad. Wireless communication enables remote patient's chart access, examination in real time for adequate preparation of hospital's team. Pocket-sized ultrasound can be important in diagnosing emergency patients, giving proper treatment, thus decreasing the risk of fatality. Medical apps for smartphones are intended for doctors, med students and patients. These innovations could bring economic, social and medical benefits for patients, doctors and government, but it seems like these are not given enough attention in Serbia.

1.1. Facts

Medical diagnostics has significantly improved over the years e.g. in determining ischemic and nonischemic heart condition. Despite that some patients are discharged without correct diagnosis, thus increasing the probability of fatal outcome. According to American Heart Association in 2010, between 2% and 8% of patients with acute coronary syndrome are discharged without the accurate diagnosis, thus increasing mortality rate [1]. Every patient will undergo physical examination and medical history analysis; other tests (e.g. blood analysis, MRI, CT, ECG) depend on cardiologist's opinion [2] and are necessary for confirming/rejecting diagnosis. Their complexity takes time; they are not always available when needed. Proper, in-time therapy is crucial, but it depends of accurate diagnosis. If diagnosis is incorrect, prescribed therapy can be unnecessary, even harmful [1].

Miniaturization [3] and wireless technology have an important role in medicine, which can be seen e.g. in usage of pocket-sized ultrasound devices as diagnostic tool [4]. Serbia does not have access to this device (Vscan Dual Probe more accurately), because it is not available for sale here. These problems raise following questions: Q1: Can PSUD help in more precise diagnostics?; Q2: Does PSUD decrease time from admittance to precise diagnosis?; Q3: Can PSUD always give precise diagnosis?; Q4: Are PSUD's performances better than traditional ultrasound device?; Q5: What are the experiences of PSUD users? [4]; Q6: How do patients, clinicians and government benefit from PSUD usage?; Q7: How can Serbia benefit from this unavailable device? and Q8: How to ensure the implementation of the PSUD, considering the obstacle Serbia has?

1.2. Objectives of Research

The first objective is to analyse, through results of different authors, Vscan pocket-sized ultrasound device, its advantages and disadvantages and its influence on early accurate diagnosis and treatment, which lead to patients' recovery and avoidance of fatal outcomes, and elimination of unnecessary costs because of unneeded tests and procedures. The second is to eliminate the problem with unavailability of Vscan Dual Probe by suggesting alternative solution, with lower costs.

1.3. Dilemmas with Pocket-sized Ultrasound Devices

PSUD are being used by medical staff, but as a relatively new and costly technology, they bring certain dilemmas (D_n), which the authors wish to investigate. PSUD have small dimensions, which means that they lack certain characteristics. D_1: Can they still provide faster and more precise diagnosing to the users?; D_2: Can their usage help avoiding unnecessary tests?"; D_3: Can they decreases time for correct diagnosis and administration of adequate therapy and thus decrease probability of fatal outcome?; D_4: Can they substitute traditional, expensive ultrasound machines? and D_5: Can they be used by other medical personnel without experience after proper training? [4].

1.4. Defining Key Terms

Miniaturization is the tendency of making robust medical devices smaller, thus enabling its mobility, but it is important to retain, if not improve, technical capabilities. *Pocket-sized ultrasound device* (PSUD) is slightly bigger than smartphone, intended for focused visualization of a patient's body. *Med apps* for smartphones, intended for medical workers, med students and patients, are used for health monitoring, prescribing treatment and as an assistant in diagnostics.

2. Discussion

Pocket-sized ultrasound devices are an important medical tool, because they allow doctors to see inside a patient, thus having a great influence on determining whether a procedure will be performed or not. This is a very significant advantage, because it saves doctor's time (unnecessary procedures are avoided), has a positive influence on patient's recovery (less procedures mean less wounds and faster recovery), thus having an impact on lower costs which would be generated with an unnecessary procedure and longer recovery time.

2.1. Research Plan

General Research Plan has 3 parts. The first discusses problems, objectives, dilemmas and definitions. The second presents different attitudes towards the usage of PSUD as diagnostic tools. The third part interprets the results of doctors using PSUD; gives answers to the questions, discusses dilemmas and suggests how to modernize MedIS in Serbia. Scientific database Mendeley and the Internet were used for conducting research – data gathering was based on articles and clinical trials relevant for this area of expertise.

2.2. Vscan

Vscan is a pocket-sized ultrasound device (PSUD) produced by „General Electric Healthcare Unit". It is general purpose diagnostic ultrasound imaging system for use by qualified and trained health care professionals enabling visualization and measurement of anatomical structures and fluid. The dual headed probe integrates both linear and phased array transducers that allows for a wide range of clinical applications. Portability and simplified user interface enables integration into examination and training sessions. The information can be used for basic/focused assessments and adjunctively with other medical data for clinical diagnosis purposes during routine, periodic monitoring, triage assessments for adults and paediatric and for procedural guidance. Complemented with gateway software, it enables image management at customers' PC, a Vscan web portal, including online access to product and clinical information. Display has 3.5" with 240x320 resolution. Probe's dimensions are 129x39x28 mm. Total weight (unit and

probe) is 436 grams. Vscan has diverse clinical applications (e.g. renal, abdominal, cardiac, obstetrics and gynaecology, fetal, urology) [5]. It is not available in Serbia.

2.3. Effects of Pocket-size Ultrasound Device

Galderisi, M. et al. in [3] assessed diagnostic power of PSUD. Examinations were conducted by experienced doctors and interns. Physical examination discovered cardiac abnormality in 38,2%, while combining physical examination with Vscan detected abnormality in 69,7% of cases. This confirms the importance of Vscan in examination and diagnosis. The study has shown result concordance between Vscan and Doppler of 0,67. However, the concordance was higher with experienced doctors – 0,84, while the interns, who had previous training, had much lower concordance of 0.58. This means that the experience of medical worker is very important during examination, when analysing the results achieved by using Vscan and Doppler.

Frederiksen et al. in [6], emphasize the huge potential of PSUD, because they enable examinations in real time and provide help in accurate diagnosis next to the patient. This is important because waiting for care jeopardizes patients' lives. PSUD can be used in different areas of medicine and therefore can be expected of it to become universal examination tool. When admitted, depending of the symptoms, patients will be examined using PSUD, thus improving diagnostic accuracy and prescribing the right treatment.

De Backer et al. in [7] see usefulness of PSUDs in emergency case because of 2D images and colour flow, but underline that because of their size, lack certain characteristics classical ultrasound machines have, like Doppler function. Despite that, they are ready for focused echocardiography in emergencies and ambulance.

Mancuso et al. in [1] compared diagnosis after medical history analysis, physical examination and electrocardiogram with diagnosis after Vscan examination, which was done by experienced cardiologist. Patients complained on: chest pain, dyspnea, palpitations, dizziness/hypotension, and lower limb edema. First diagnosis: acute coronary syndrome, heart failure, arrhythmia, pulmonary embolism, cardiac effusion/tamponade, shock of undetermined etiology and aortic dissection. Second diagnosis: in 90 cases, first diagnosis was inconclusive. After Vscan, only 28 initial diagnosis were confirmed; 17 were rejected and new diagnosis was determined, thus implying that wrong treatment could have been applied, as well as fatal outcomes; 45 remaining were inconclusive and had to be tested.

Dr Alan Garner [8] from Care Flight emphasizes the usage of ultrasound for diagnostic as well as its robustness, which made it impractical for field interventions. But miniaturization has enabled the ultrasound to become vital part of medical equipment for field caring. E.g. it is important to determine whether the lung is up, or is it necessary to make a hole in the chest, for the leaked air to escape and allow the presumed collapsed lung to re-expand. Before PSUD, the team had to do this, although sometimes it would turn out to be unnecessary. Now, they can see and know whether a procedure should be performed or not. Examination with stethoscope is difficult, while other teams are freeing a victim and in helicopter it is impossible. Vscan provides high-resolution images using the high-frequency linear probe which gives clear views of the function of organs such as lungs that are close to the skin's surface; but without changing probes, the user can switch to the phased-array transducer, which offers lower

frequency, more deeply penetrating views of abdominal or thoracic cavities, pelvic disruption or heart motion.

Testuz et al. in [9] tested Vscan with one probe in acute cardiac care and results were compared with echocardiograph. The agreement for the left ventricular systolic function and pericardial effusion was high, 0.89 and 0.81, respectively. The agreement was good or moderate for evaluating the aortic, mitral, and tricuspid valve function and the left ventricular size: 0.55–0.66. Visualization of the Vscan images in full-screen format on a PC did not in general confer added value. Doctor performing ultrasound examination has to be experienced, whether using PSUD or echocardiograph, especially in emergency room, where conditions are stressful and wrong diagnosis could have fatal outcome.

2.4. Alternative for Vscan Dual Probe

Vscan Dual Probe is not available for sale in Serbia. One solution would be to contact the GE Healthcare Unit's representatives and negotiate the launch of the device on Serbian medical devices market. In case of failure, there is another alternative.

Smartphones are used by medical staff, med students and patients for health monitoring, prescribing treatment and as diagnostics assistants. The idea is to search for the Internet to find a suitable app for ultrasound or simply form a team of experts to design the application for smartphones, which would have to be compatible for Windows, Android and iOS. That app would be installed on the smartphone and the probe would then be connected to the phone via USB port.

Resolution of the smartphone should be equal or higher than the resolution of the transducer. There are different types of probes: *abdominal* for bladder, aorta, liver, kidney, gall bladder, OB; *peripheral imaging*: vascular access, nerve blocks, testes, intraoperative, musculoskeletal; *endocavity*: transrectal, transvaginal. Each could be connected via USB to netbook, laptop, desktop [11] and smartphone. Since smartphones have micro USB, the micro USB to USB cable should be used to connect the phone on one end to the transducer on the other. The manufacturers of these probes also provide applications to match the transducer, which would be installed on the smartphone from the App Store free of charge. Every time a doctor needs an ultrasound, s/he would simply connect the transducer to the smartphone and perform examination.

Another solution is getting rid of all those cables, because of ease of handling the transducer as well as for medical reasons (sterilization of the cable and the probe e.g. for catheterization) by deploying wireless technology to connect smartphone with transducer, thus enabling comfort, as well as patient safety during examination where ever it is performed.

The medical device has a transducer, an echocardiograph and a Wi-Fi antenna all in one [12]. An application from manufacturer must be installed on the smartphone from the App Store free of charge. The transducer has the ability to connect via Wi-Fi with the smartphone. It has an antenna, which sends the signal to the smartphone. A doctor goes to settings and enables Wi-Fi connection on the smartphone. Then the transducer is turned on and the Wi-Fi signal is sent. The smartphone registers the transducer, and the doctor enters the number on the back of the probe, thus establishing the connection. After that, the doctor launches the application on the smartphone, applies the gel,

unfreezes the screen (the initial app mode is freeze) and performs the examination. In case of guidance procedures e.g. catheterization the transducer is put in sterile bag.

The connection between the smartphone and the probe does not need the Internet. This would enable doctors to perform ultrasound during general examination, in emergency rooms and on the way to the hospital, whether they are in urban or rural areas. They could also save the examinations in video and audio format, and send it to colleagues for consultations and ask second opinion, in real time. Of course, the Internet connection would be necessary in this scenario. The importance of this alternative is that it saves money, because the Vscan costs a few thousand dollars and the set of wireless transducers from approximately 2.000 to 3.000 dollars, depending on the set.

2.5. Vscan Dual Probe has a flaw!

Different authors have pointed out different problems with the PSUD. However, there is one very important flaw with Vscan Dual Probe Device. Since doctors also use it in ambulance and aircrafts, it needs a characteristic which would allow doctors to send data to hospitals and consult their colleagues in real time, thus, if possible, administering proper therapy, and save time and money on the way to a hospital. Also, it would be necessary to provide Vscan with software which would protect patients' data during transfer.

3. Conclusion

After comparing primary and secondary examination, Vscan had better results. Cardiologist confirmed 28 and changed 17 initial diagnoses [1], so patients were prescribed different, adequate therapy and/or surgical procedure on time.

E.g. 2 out of 3 patients, where cardiac tamponade was confirmed [1] had to undergo pericardiocentesis, so that the liquid pressuring the heart (myocardium) could be drained from pericardial sac. Otherwise, the liquid would continue pressuring the heart, causing problems with its functioning, thus leading to fatal outcome.

1 case with heart failure was actually severe aortic stenosis [1], which meant that the patient had to insert artificial valve, to normalize blood flow and save heart from weakening.

Classical examination could not give certain diagnosis in 90 cases and Vscan in 45, thus implying that doctors achieve better results when using Vscan [1]. This has impact on patients and doctors, because Vscan decreases the number of patients who are sent to specialized, time consuming, expensive tests, thus decreasing the waiting lists and enabling early adequate treatment which makes a difference between life and death, as well as cutting costs significantly.

Although Vscan gives good quality 2D images [1] and there is no difference between looking an image on PSUD and computer [9], as well as colour flow mapping and some linear measurements, the devices does not measure the area and volume and does not have continuous or pulsed Doppler [1].

Based on opinions of different authors, PSUDs are important in critical situations, because they allow examination in Point-Of-Care (POC), relatively quick accurate

diagnosis, avoidance of unnecessary test, thus decreasing time spent from admission till prescription of proper therapy. This decreases the number of fatal outcomes. Therefore, despite its size, the PSUD can make a difference in diagnosing and treatment, especially in emergency situations, thus resulting in positive outcomes (D_1, D_2 and D_3).

However, diagnostic capabilities are not exceptional. It has some technical limitations and doctors can be uncertain when determining a patient's illness, and will require additional tests for setting proper diagnosis. De Backer et al. in [7] and Mancuso et al. in [1] emphasize that certain measurements are not possible, so traditional, expensive ultrasound machines will still play an important role in diagnostics (D_4), although with continuous improvement of PSUD, substitution of classical ultrasound machines can be expected, especially when analysing the improvement, capabilities and size of personal computers through time.

The research in [3] proved a significant difference in concordance of Vscan and Doppler examinations between, experienced doctors (0,84) and trained interns (0,58) and research in [9] underlined the importance of experience doctors should have when performing ultrasound examination, whether using a PSUD or echocardiograph, especially in emergency room, where conditions are stressful and wrong diagnosis could have fatal outcome. Therefore, PSUD can only be used by experienced medical personnel (D_5).

Serbia can either contact the company or use what it has: smartphones with developing or downloading software for using probe via USB connection. Another solution is to buy wireless transducers, which are much more affordable and apps for them are free of charge.

The authors criticize the inability of the Vscan Dual probe to send patient data to hospitals on the way there, because there is a technology which can be deployed and the concept of telemedicine would provide real time communication related to images obtained in that way. It is expected that GE Healthcare Unit will do something about this in the future.

Improvement of pocket-sized ultrasound devices' characteristics can be expected in the future: higher-resolution image quality, Doppler function, higher diagnostic accuracy, more powerful battery. It is expected that examinations will be performed by trained medical staff, not just experienced doctors. This will allow PSUD usage not just in an ambulance and emergency room, but in everyday check-ups and remote areas, with the primary goal of saving peoples' lives, thus decreasing healthcare costs, such as unnecessary tests, procedures, patient's recovery and absence of work.

References

1. Mancuso,F., Siqueira,V., Moisés,V. et al.: Focused cardiac ultrasound using a pocket-size device in the emergency room. Arquivos brasileiros de cardiologia. Vol.103, No.6, 530-537. (2014)
2. Mayo Clinic Staff. Blood Tests for Heart Disease (7 December 2016) [Online].Available: http://www.mayoclinic.org/diseases-conditions/heart-disease/in-depth/heart-disease/art-20049357?pg=1 (26 August 2017)
3. Galderisi, M. et al.: Improved cardiovascular diagnostic accuracy by pocket size imaging device in non-cardiologic outpatients: the NaUSiCa (Naples Ultrasound Stethoscope in Cardiology) study. Cardiovascular Ultrasound.(2010)

4. Gračić, S.: Medicina u pokretu - primena robotskih hirurških sistema, „džepnih“ ultrazvučnih uređaja, „pametnih“ telefona, transportnih sistema i drugih dostignuća iz oblasti telemedicine zasnovanih na bežičnim tehnologijama i biometrici u Srbiji i inostranstvu. Master rad, Univerzitet u Novom Sadu. Ekonomski fakultet Subotica (2016).
5. GE Healthcare Unit: Vscan with Dual Probe. Data Sheet. (2014)
6. Frederiksen C.A., Juhl-Olsen, P., Sloth, E.: Advances in imaging: ultrasound in every physician's pocket. Expert Opinion on Medical Diagnostics, Vol.6, No.3, 167-170. (2012)
7. De Backer D, Fagnoul D.: Pocket ultrasound devices for focused echocardiography. Critical Care. Vol.16, No.3, 134. (2012)
8. GE Reports Australia, New Zeland & PNG. Look! Up in the sky! It's ultralight, lifesaving ultrasound! (24 July 2015). [Online].Available:http://gereports.com.au/post/24-07-2015/look-up-in-the-sky-it-s-ultralight-lifesaving-ultrasound (26 August 2017)
9. Testuz, A., Muller, H., Keller, P.F., Meyer, P., Stampfli, T., Sekoranja, L., Vuille, C., Burri, H.: Diagnostic accuracy of pocket-size handheld echocardiographs used by cardiologists in the acute care setting. European Heart Journal – Cardiovascular Imaging, Vol.14, 38–42. (2013)
10. Mayo Clinic Staff. Heart Disease. (2 August 2017). [Online].Available: http://www.mayoclinic.org/diseases-conditions/heart-disease/diagnosis-treatment/diagnosis/dxc-20341570 (26 August 2017)
11. National Ultrasound. Interson USB Probes. (n. d.). [Online].Available: https://www.nationalultrasound.com/ultrasound-manufacturers/interson-seemore-usb-probes/
12. Moncada, M.: Ecografo de Mano Ecopalm Wifi. (1 June 2016). [Online].Available: https://www.youtube.com/watch?v=HODYPvT7xUU (20 September 2017)

Development of Low-Cost Wireless Sensor Network for Solid Waste Management

Goran Mateski[1], Ramona Markoska[1], and Aleksandar Markoski[1]

[1] Faculty of information and communication technologies, Partizanska bb, 7000 Bitola, Macedonia
goran.mate@yahoo.com, ramona.markoska@fikt.edu.mk, aleksandar.markoski@fikt.edu.mk

Abstract. Waste generation is one of the major problems in modern societies, not only for the quantitative of waste generated, but also for the increasing complexity of process of waste collection. Main objective of the different methods for collecting the waste routing optimization to minimize the total route cost. Algorithms which are solving Vehicle Routing Problem depends on real-time data which can be provided only by ad sensors. A low-cost sensor node and sensor network for waste detection are developed and presented in this paper. Sensor node is based on ultrasound sensor and Arduino open-source hardware and software. Measured data are collected and stored in Thing Speak cloud server, and they can be used for visualization, analysis, and as input for algorithms solving Waste Collection Vehicle Routing Problem.

Keywords: solid waste detection, ultrasonic, sensor network.

1. Introduction

Solid Waste Management is one of the most complex problems in modern societies, not only for the quantitative rise of the amount of waste generated, but also for the increasing complexity of some products and components. Waste collection is a very complex logistics operation, and how to collect waste in an efficient way is an area that needs to be improved.

Solid Waste Management System architecture is presented on Fig. 1. Components of the system are: waste sensors, RFID tags, communication modules, GPS units, database servers, software for integration and visualization of all collected data, and finally software which solves the most important problem: routing optimization to minimize the total route cost. This problem in literature [1] is defined as Waste Collection Vehicle Routing Problem - WCVRP.

According to [2] identification of a route for Municipality Solid Waste collection trucks is critical because from the total amount of money spent for collection, transportation and disposal of solid waste, approximately 60–80% is spent on the collection phase. Significant savings of money, fuel and time can be obtained using the combination of modern technologies providing the necessary data (sensor network, databases) and algorithms for intelligent routing (Fig. 2).

Fig. 1. Architecture of the Solid Waste Management System

In addition, by lowering the number of collections the noise of garbage trucks will be decreased significantly and the emission from trucks in the atmosphere will be lowered because the fuel consumption will decrease.

Fig. 2. Static vs Intelligent routes based on Sensor data

The main component of the Solid Waste Management System is a sensor node capable to identify the level of the waste of so called “smart bin” shown on Fig. 3.

Fig. 3. Concept of "smart bin" with installed ultrasonic sensor

In the literature, many methods and principles of waste detection have been used [3], [4], [5]. Also there are several commercial solutions at the market. Availability and very low price of the electronics, controllers, sensors, and communication modules based on Arduino open-source hardware, gives us opportunity to develop a low-cost sensor nodes and sensor network which will be used to provide necessary data for development of algorithms and software for WCVR Problem solving.

2. Development of Sensor Node

The main component of the system is ultrasonic module HC-SR04 used for non-contact measuring from 2cm up to 400cm with 3mm precision. It consists of ultrasonic transmitter, receiver and circuit. The function is based on sharing 8 signals of 40 KHz and checking if there is a returning pulse. After that if we have returning pulse while the I/O port is opened, we get the period from sending until receiving of the signal (Fig. 4).

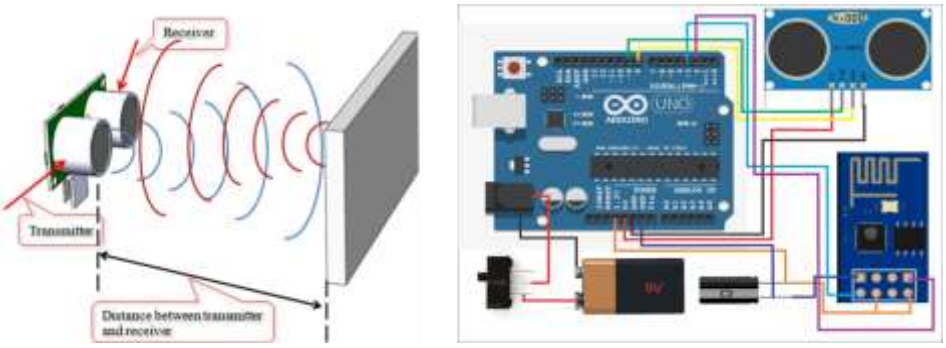


Fig. 4. Principle of measurement and components of the sensor node

The sensor node uses microcontroller Arduino Uno based on AT Mega 3328, ultrasonic sensor type HS-SR04, module for wireless communication WIFI ESP8266, battery 9V, capacitor 2200 μF and slider switcher with two conditions. (Fig. 4).

ESP8266 is used for wireless networking, and enables usage of applications or transfer all of the wireless network functions onto application space. Its usefulness is proven by the amount of data that it can process and collect. This system include fast reaction, adaptable radio waves while low expenses, processing the signal in advance and capability for networking with other technologies. Main element in the sensor node is the microcontroller, Arduino Uno onto which are connected other elements. The ultrasonic sensor HC-SR04 includes 4 pins from which, 2 pins are used for powering and 2 of them are used for communication with the I/O ports of the microcontroller.

Working principle is based on 5V voltage which enables direct connection with Vcc pin and the pin for powering from 5V out of microcontroller. Grounding pin (GND) is connected to one of the two pins from the microcontroller. The rest of the pins from the ultrasonic sensor are connected to the I/O digital ports of the microcontroller, with consideration of the programming code. While wireless module for communication is into working mode, it uses higher level of powering, around 50mA which may cause the module to be prevented from function. In order to avoid such a situation, capacitor from 2200μF is used. Pins for powering (Ch_PD) and (+3v3) from the module, are connected to the positive pin from the capacitor while pin GND is connected to the pin for powering from 3.3V of the microcontroller, the same goes with the negative pin of the capacitor.

In order to establish wireless communication, microcontroller uses two pins, Rx and Tx, with the usage of different locations of connection while programming and connecting. Powering the node is established by battery from 9V which includes two contacts (positive and negative). For changing state of work, slider switch is used in two states and 3 pins for connection. All elements are packed in practical box, which must be easy for packing and unpacking so that whenever there is an element out of function, it could be replaced immediately. In order to be programmed and configured, the microcontroller together with additional elements, uses the software Arduino IDE 1.6.7 version.

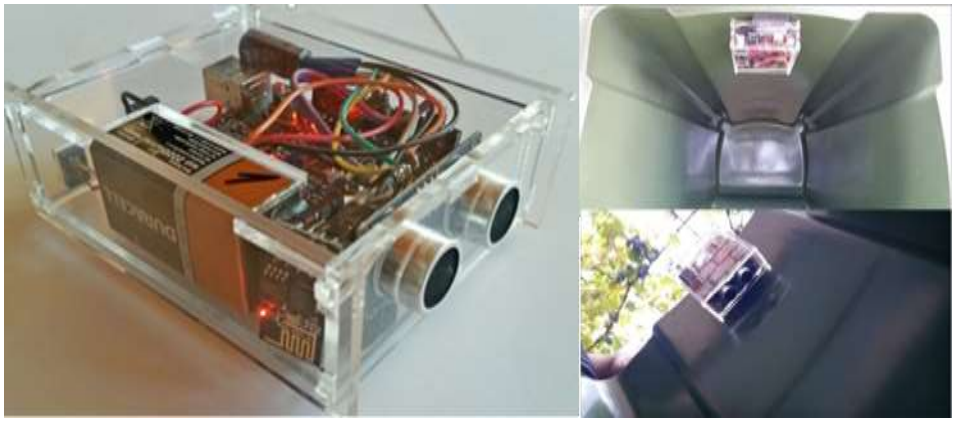


Fig. 5. Complete Sensor Node and placement in “smart” bin

Based on the ultrasonic sensor principle of work, the node is positioned near to the top edge of the garbage container. Ultrasonic sensor does the measuring with sending radio wave which is declined of some obstacle inside the container and is turned back to

the receiver. In the following way, the calculation of speed, time and density, determines the value of the garbage which is contained into the container.

The process starts with turning ON the switcher on the node, after that follows the process of measuring the distance from ultrasonic sensor until the bottom of the container. When turned on, the container should be empty in order to record the maximum distance (calculating the space between the minimum and maximum level in the garbage container). When the distance is recorded, the controller is activating the module for wireless communication who is sending information to the repeater. When connected to the router, measured data is sent and the connection is interrupted. If the router is not available, the process is repeated until the connection is not established.

According to the program, the time interval of measuring and sending data to the repeater can be lower than 5 seconds and it is repeated without interruption. Server is available to display eight of the sensors data in eight fields adapted for that data. With declaration of the address in program, data is displayed on the particular field from channel on the server in every 15 seconds. In real application time interval should be much bigger, in order of several minutes, which will lower power consumption and it will increase battery life.

3. Architecture of the system

Because main aspect of the node development is providing data for development of the algorithms and software for WCVRP, we used very simple network architecture for connecting the sensor nodes to the system using Wi-Fi connection with ESP8266 shown on Fig. 6. Instead of building complete server and database solution, measured data are sent to ThingSpeak cloud (https://thingspeak.com/).

Fig. 6. Integrated System for Waste Management and Control

The ThingSpeek server has capability to collect data at real time, visualize it in collected form out of the data table, produce plugins and applications for using web-

services, social networks and other platforms. ThingSpeak enables usage of a big variety of different visualization types and activities that may be inputted in the page with graphs by special plugins. There are two types of plugins, ones are for visualization and analytics and another ones are based on activities. Analytical data is used for precise visualization based on the type of the data. The activity plugins are used for control and representation of changing activities.

In order to be used, it is obligatory to have profile and log into the official page www.thingspeak.com where the channel will be created. With registering on the channel, the API code is also created as unique for that profile. The profile could contain more channels and the data could be represented either public or private (only for the user). API codes are used while programming microcontrollers and its components. There are two API codes, API code for writing is automatically generated by the channel, opposite to the API code for reading which must be generated by the user. In that way the communication is established and the data may be transferred.

4. Testing of the System

In order to test the system three Sensor Nodes are deployed on the city of Ohrid, Macedonia, on location shown on Fig. 7., with measured data and graphs from ThingSpeek are shown on Fig. 8.

According to our experience, there are a lot of places where you can position the wireless sensor node, but the best place according to our research is on the top of the container. This principle of working has many disadvantages. Basically the container may be left opened sometimes because of inattention which may result in inaccurate results, also vandalism which is often present in the streets may cause problems from different kinds with the node inside. But there is a solution for these problems too, like positioning sensors for heat, vibrations etc. They may register and send these data to particular
like police or fire station in order to report vandalism or waste burning.

Fig. 7. Locations of tested "Smart" bins

For more precise display and review of the data in the system for control and maintenance of the garbage, Google plug-in is used, which displays measured data in

form of gauge in different colors (Fig. 9.) which shows the current level of waste in the bin. The colors are adapted to the level of garbage in the container

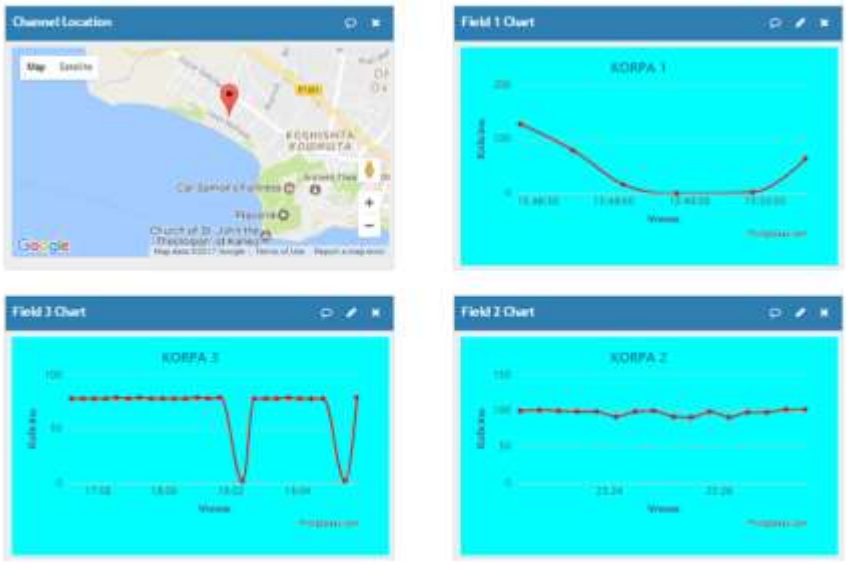


Fig. 8. Graphs from measured data on ThingSpeek

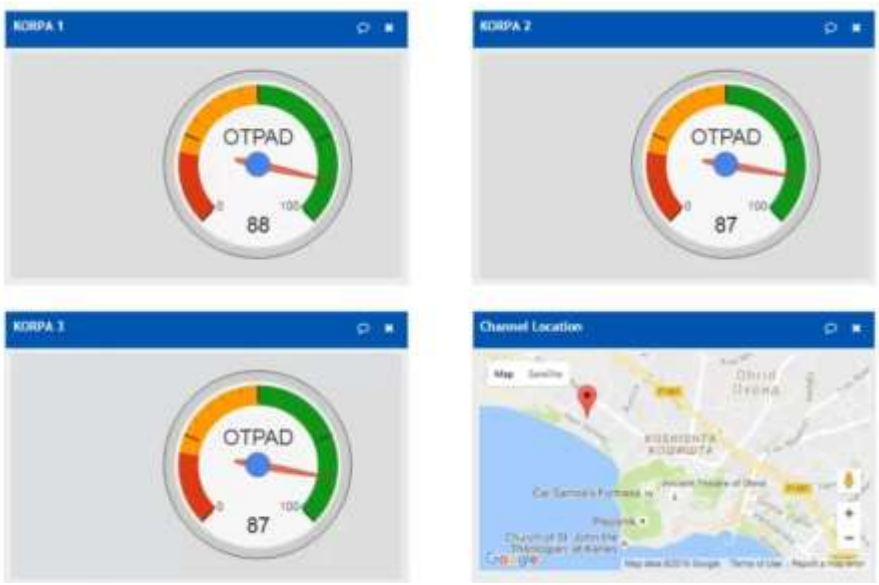


Fig. 9. Current waste level

ThingSpeak Server is capable to collected 8000 measurements on single channel and that data can be saved in form of CSV format. It allows data to be inputted into the software for analysis and collecting data for further processing in WCVRP modelling.

If it is necessary, sensor node can be updated with additional sensors and modules in order to be capable to send and receive more data and perform more precise measuring. In order to enable the node to perform activities outside the community where wireless networks are not available, system could be adapted to communicate by GSM/GPRS module which may perform the same function on wider area. For longer performance of the battery, life of the battery may be prolonged using solar charging or batteries with bigger capacity.

5. Conclusions

This paper focusses on development of low-cost wireless sensor network for solid waste management. Complete solution based on Arduino open source hardware, sensors and communication modules is presented. The solution is simple, can be realized with minimum investment per sensor node, but still effective for various applications. In this version, all data are stored in ThinSpeak cloud which can be also used for analyses and visualization. System was tested on-field and first result shows that this concept can be improved and widely used not just for research purposes, but also in real practical applications as a part of the Solid Waste Management System.

References

1. H. Han, E. Ponce-Cueto: Waste Collection Vehicle Routing Problem: Literature Review, Promet – Traffic&Transportation, Vol. 27, No. 4, 345-358, (2015).
2. Karadimas N.V., Papatzelou K., Loumos V.G. Genetic Algorithms for Municipal Solid Waste Collection and Routing Optimization. In: Boukis C., Pnevmatikakis A., Polymenakos L. (eds) Artificial Intelligence and Innovations 2007: from Theory to Applications. AIAI 2007. IFIP The International Federation for Information Processing, vol 247. Springer, Boston, MA, (2007).
3. Md. Abdulla Al Mamun, M. A. Hannan , Aini Hussain , Hassan Basri, Wireless Sensor Network Prototype for Solid Waste Bin Monitoring with Energy Efficient Sensing Algorithm, Proceedings of the 2013 IEEE 16th International Conference on Computational Science and Engineering, p.382-387, December 03-05, (2013).
4. S. Longhi et al., Solid Waste Management Architecture Using Wireless Sensor Network Technology, in 5th International Conference on New Technologies, Mobility and Security (NTMS), pp. 1-5, (2012).
5. V.Catania, D.Ventura, An approch for monitoring and smart planning of urban solid waste management using smart-M3 platform, 15th Conference of Open Innovations Association FRUCT and 3rd Regional Seminar on e-Tourism, (2014).

A General Overview of the Computer-Assisted Translation Tools

Nastasja Deretić[1], Olga Deretić[2]

[1] Proverbum d.o.o., Hajduk Veljkova 11/IV,
21000 Novi Sad, Republika Srbija
nastasja.deretic@gmail.com
[2] Technical College of Applied Sciences
in Zrenjanin, Đorđa Stratimirovića 23,
23000 Zrenjanin, Republika Srbija
olga.deretic@gmail.com

Abstract. Nowadays, technology is integrated into all aspects of human labor; it increases their productivity and contributes to the quality of the result. Translation is not an exception and there has been an expansion of a large number of computer programs which can facilitate a translator's job immensely. Different computer-assisted translation tools (CAT tools) focus on different aspects of the translation process (terminology, consistency, spelling, grammar, etc.), but they basically all serve the same purpose: to help the translator to provide the translation of the highest possible quality in as little time as possible. This paper provides an overview of the basic CAT tools which are presently used by most professional translators, as well as of their crucial features and principles they are based on.

Keywords: computer-assisted translation tools, translation, information technologies.

1. Introduction

The development of technology regularly leads to a progress in various human activities. The implementation of modern technological solutions contributes to both acceleration and efficiency of all activities performed within a certain job, as well as to improving the quality of the end result, i.e. product or service. The translation business is certainly not an exception to this rule. Nowadays, translators have almost completely stopped using pens and paper in their job, a typewriter has become a thing of the past, and every day there are fewer and fewer professional translators whose only tools are regular text editors, electronic and printed dictionaries, as well as their own knowledge and experience, as vast as they may be. The last decades have welcomed computer-assisted translation tools, which are considered technology's greatest contribution to translation business.

2. Computer-assisted translation tools

Computer-assisted translation tools or CAT tools are computer programs designed to facilitate the translation process, reduce the time necessary for the translation of a given text, as well as to ensure the consistency and the quality of the target text, i.e. translation. At this point, however, it is very important to emphasize the difference between computer-assisted translation and machine translation. In machine translation, the computer is expected to deliver a complete translation of a text (or its part) from one language to another; on the other hand, computer-assisted translation involves an interactive process between the computer and the translator. In other words, the translator is still responsible for the final product, and he/she may never entirely rely on the software. The translator still has to use his/her knowledge and experience to ensure the highest possible quality of the translation.

3. Types of computer-assisted translation tools

There are many different computer-assisted translation tools and each of them simplifies the translator's job in a certain way and improves certain aspects of the translation. The basic CAT tools which every professional translator should know how to use are: segmentation tools and translation memories, linguistic search engines, term bases, alignment tools, machine translation tools, and quality assurance (QA) tools.

3.1. Segmentation tools and translation memories

Most computer-assisted translation tools are based on the segmentation of texts. These text segments usually consist of no more than two sentences (some of these tools can be additionally adjusted so the translator can prevent the tool from perceiving the full stop within an abbreviated form of a word as the end of the sentence, thus preventing the splitting of a single sentence in two segments). In this way, the translator is provided with a much clearer layout of the text since the sentences of the source text are aligned with the fields in which the translated sentences are typed in.

According to Champollion [1], even when used independently (without the translation memory), this kind of segmentation is very beneficial. Above all, the translator's work requires less physical effort: he/she does not have to constantly look down from the screen onto the paper and vice versa; he/she does not have to keep opening and closing different text editor windows; finally, he/she does not have to keep searching for the part of the source text that he/she is supposed to translate next. Thanks to the segmentation tools, the eye strain is much more reduced, which is extremely favorable, both in the short run, since fewer breaks are required, and in the long run, since the risk of sight problems is considerably decreased. These tools also reduce the possibility of a lapse of concentration and mental exhaustion.

Another advantage of these tools is that they eliminate the possibility of skipping an entire sentence or even a paragraph in the course of the abovementioned search for the next sentence to be translated from the source text (this major mistake may happen in a text containing a large number of similar sentences).

Finally, the segmentation tools automatically create the text layout corresponding to the source text layout, so the translator does not have to waste time formatting the final translation document.

Segmentation tools are rarely used independently; they are usually used in combination with a translation memory (TM). Every time a segment is translated, the computer saves the translation as a translation unit (TU) within the translation memory. Each TU records the source and the target segment, date and time of creation, source and target language, as well as the ID of the translator who entered the translation. In addition, the frequency of the use of a particular translation unit is also recorded, which determines the value of the translation of the segment in question.

When translating materials such as legal or technical documents, which often include a large number of repetitions, using a TM can save hours of work since the translator does not need to keep typing in the translation – he/she only needs to check it and confirm it [3]. When it uses a translation memory, the computer does not only search for exact matches, but also partial, "fuzzy" ones. The degree of similarity between segments is expressed in percentages: 101% means that the segments are exactly the same both in terms of content and format; 100% refers to an exact match of texts. The suggestions with values lower than 100% are considered fuzzy and should be used with caution; in these cases the differences between the current source segment and the source segment derived from the TM are usually colored or highlighted [1].

When it comes to the origin of translation memories, each translator can create his/her own personal TM. However, it is usually the case that the TM is provided by the translation agency of which the translator is an employee, or even by the client himself/herself. Moreover, a TM for a certain area of translation can be purchased or shared with other translators.

The most popular tools of this type are *SDL Trados*, *Wordfast*, *memoQ* (Figure 1), *Across*, *DéjàVu*, etc.

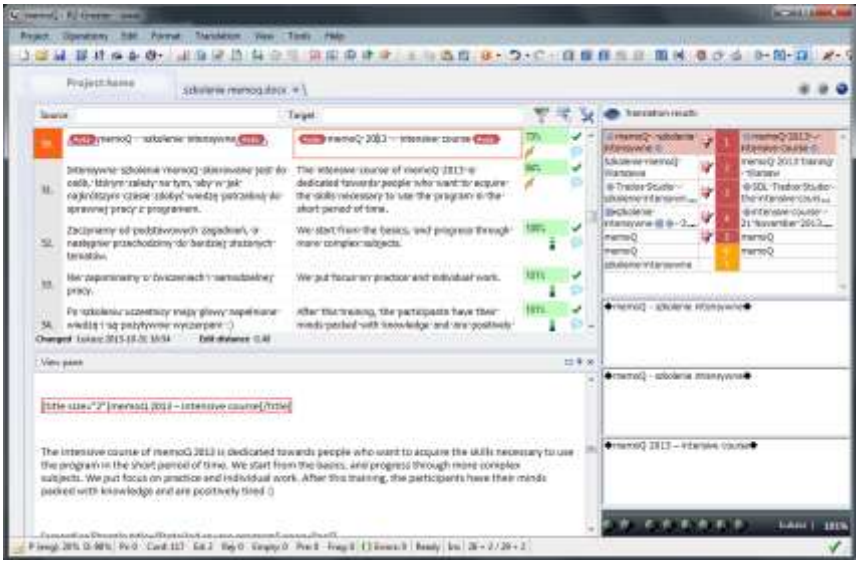


Figure 1: memoQ – Segmentation tool with TM [4],

3.2. Linguistic search engines

Linguistic search engines operate in a similar manner as regular search engines such as *Google*. However, while *Google* search offers different websites, a linguistic search engine browses through large TM databases and locates segments of previously translated texts and the corresponding translations. In this way, the translator can gain insight into different contexts in which a particular phrase may occur and choose the translation from a context which is the same or similar to the one that he/she is dealing with. *Linguee* (Figure 2) is one of the popular search engines of this type and it can be used online completely free of charge.

Figure 2: Linguee.com – Linguistic search engine [5]

3.3. Term bases

A translator may be asked to translate many different types of texts: novels, poems, contracts, directives, court decisions, scientific papers, etc. When it comes to, for example, scientific texts, they can refer to any branch of any scientific field. However, every translator cannot be an expert in all of these fields, and each of them has its own terminology which can be very problematic for the translator. This problem is not limited to scientific texts: almost every renowned company has its own "preferred

language", i.e. a specific terminology which is established and used in the course of their business. Not being familiar with the terminology can obstruct and slow down the translation process, even if the text is quite simple on the whole. For this reason, translators are often provided with a glossary with the terms which are to be used in the translation of a certain document. These glossaries can be created by the employees of a particular company or the scientists and scholars specialized in a particular scientific field, together with translators, and they may contain up to 10,000 entries. Therefore, a translator cannot be expected to memorize the translations of all of those terms; a translator can obviously find those terms online or in a dictionary, but it will surely take up a lot of his/her time. In addition, the translator may not even recognize that a certain word is actually a term and translate it by using words which have the same or similar meaning, but are not in accordance with the client's requests. In order to solve all of these problems, translators started using term bases.

Term bases are glossaries integrated within CAT tools. The tools containing term bases provide simple search, simple access, and, if necessary, simple modification, so that the translator can use the tool effortlessly and be sure that the provided data are reliable. In addition, a good CAT tool recognizes and highlights the terms contained within the term base in the text in order to draw the translator's attention to them [1]. During the translation process the translator may enter a new term in the term base, if the term has been approved by the client. After the translation is finished, the CAT tool warns the translator if a certain term has not been translated in accordance with the term base, thus ensuring the consistency of the translation [7].

The examples of this type of CAT tools include *SDL MultiTerm*, *LogiTerm* (Figure 3) and *Termex*.

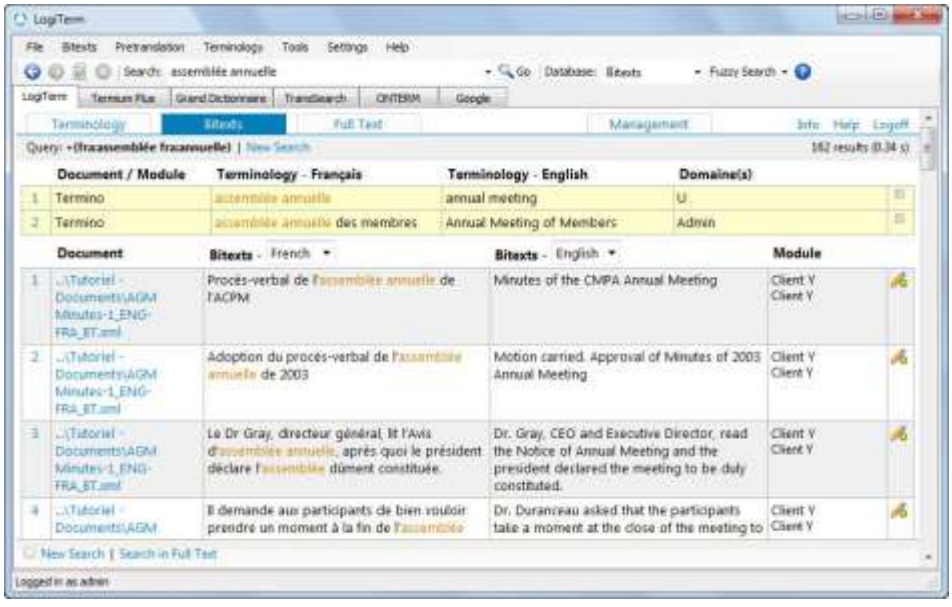


Figure 3: LogiTerm – Term base [6]

3.4. Alignment tools

The use of alignment tools is the simplest and the fastest way to create a translation memory. This program uses a source text and a translation of that text, splits them into segments, and then tries to connect the corresponding segments. The translator's task when using this tool is to check whether the segments have been connected correctly and to correct potential mistakes that the computer has made. These mistakes can occur if, for example, one sentence in the source text has been split into two sentences in the target text; however, thanks to the merging and splitting options these problems can easily be solved. After all segments have been properly connected, i.e. aligned, the resulting translation memory can be used for future translations. If a translator has not always used CAT tools when translating and therefore has not created his/her own translation memory, he/she can use this tool to include his/her previous translations into the TM. Some of the most widely used alignment tools are *Bitext2*, *Tmx Bligner*, *YouAlign* and *LF Aligner* (Figure 4).

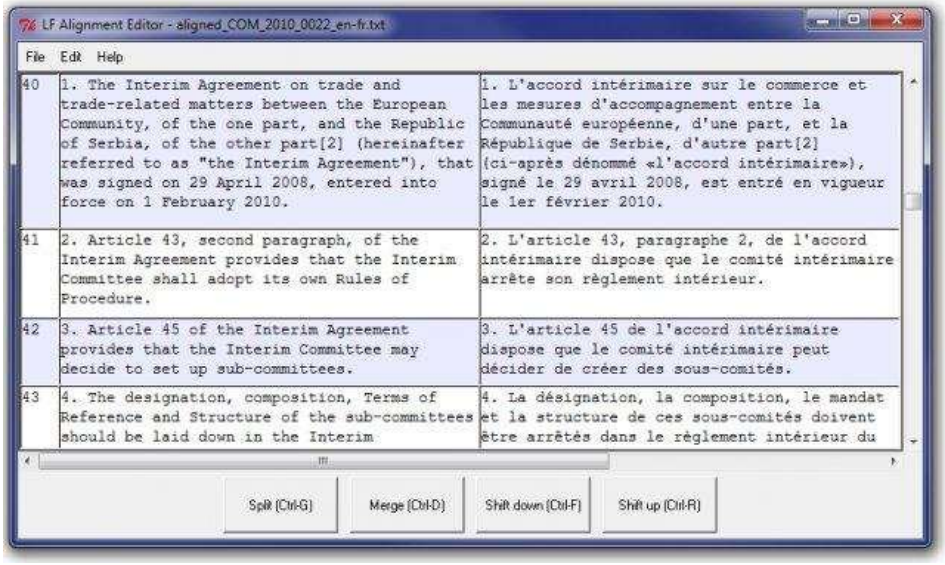


Figure 4: LF Aligner – Alignment tool [7]

3.5. Machine translation tools

Machine translation is a type of translation in which the computer uses the translation memory, term bases, and everything else at its disposal, in order to translate the text without the help of a human translator. This kind of translation is still in development, and it is widely debated whether there is a limit to that development and whether computers will ever be able to completely replace human translators. Some translators use machine translation to save time and because they find it easier to correct a poor or

incomplete translation, than to type it in completely [2]. However, this kind of translation can only provide solid results when translating very short and clear sentences, especially if they are repeated. This may lead to a conclusion that machine translation will never completely replace human translators, but it may reduce the scope of their work, and consequently their income. The most popular example of this kind of translation tool is the *Google Translate app*.

3.6. Quality assurance (QA) tools

Quality assurance tools are programs which are used to check the quality of the translation in terms of consistency, spelling, grammar, the consistency of numbers within the text, etc. Some QA tools, such as *Verifika* (Figure 5), can check all aspects of text quality. There are also certain QA tools which are focused on separate aspects:

Spell checkers – they can be used to check the spelling of the words used in the target text. They can be integrated in another CAT tool, but there are also independent programs of this type such as *Proofread*;

Grammar checkers – they can be used to check the grammar of the target text using the parameters determined for each language. They can be integrated in another CAT tool, but there are also independent programs of this type such as *Grammarly* i *Reverso.*

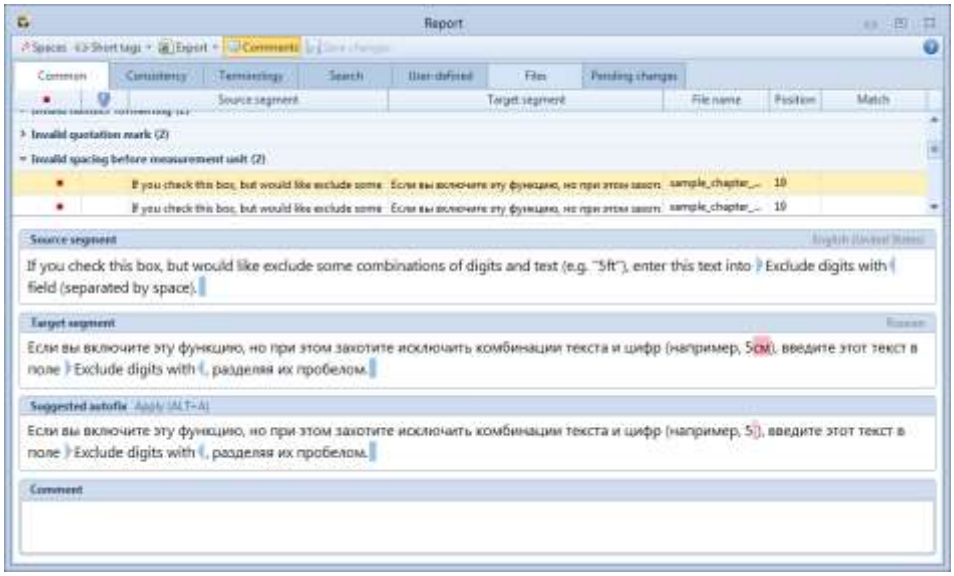


Figure 5: Verifika – QA tool [8]

4. Conclusion

Computer-assisted translation tools described in this paper are only the basis of the synthesis of translation business and information technologies, and there are also many other programs which can facilitate the translator's work. Today, when time is money and when high quality in short time is the imperative for the survival in the market, these tools are the irreplaceable assistants of every modern professional translator. CAT tools are a convincing proof that the implementation of technology raises human activity on a higher level and that it is not something to be feared, but something to be embraced and maximally used.

References

1. Champollion, Y.: Convergence in CAT: blending MT, TM, OCR & SR to boost productivity. Translating and the Computer 25. (2003)
2. Hutchins, J.: The development and use of machine translation systems and computer-based translation tools. International Journal of Translation 15.1, 5-26. (2003)
3. Sánchez, P. M.: Electronic tools for translators in the 21st century. Translation Journal 10.4, 48-54. (2006)
4. memoQ: Translation software.
   Available at: https://www.memoq.com/en/ (August 2017)
5. Linguee | Dictionary for German, French, Spanish, and more.
   Available at: http://www.linguee.com/ (August 2017)
6. LogiTerm – Terminotix.
   Available at: http://terminotix.com/index.asp?content=brand&brand=2&lang=en (August 2017)
7. LF Aligner download | SourceForge.net.
   Available at: https://sourceforge.net/projects/aligner/ (August 2017)
8. Translation Quality Assurance Tool - Verifika QA. Available at: https://e-verifika.com/ (August 2017)

IoT and Education

Željko Eremić[1], Milorad Rančić[1]

[1] Technical College of Applied Sciences in Zrenjanin, Đorđa Stratimirovića 23, 23000 Zrenjanin, Serbia
{zeljko.eremic, milorad.rancic}@vts-zr.edu.rs

Abstract. Internet of Things (IoT) is a collection of objects or devices that are interconnected in a network. Through the network they share information using appropriate standards and protocols. The goal is to create smart environments with application in various fields of everyday life. This paper will focus on the implementation of IoT in the field of education, which will be illustrated with appropriate examples. First example shows possible role of IoT in learning process. Second example shows that learning to create IoT system in not difficult process using SenseBoard hardware, Sense language and cloud infrastructure. Using mentioned equipment it is possible to make first program in only twenty minutes.

Keywords: IoT, education, smart environment.

1. Introduction

IoT (Internet of Things) is a term that has been in the focus in the recent years. For a long time, the use of computers typically required that a person who gives instructions to the computer which the computer would execute. In IoT, the focus is on that different devices are interconnected and they exchange information with one another. The result is a smart environment.

It should be emphasized that IoT is already widely present in everyday life, although we are not always aware of it. One of the areas in which IoT has been applied is education. Other applications of IoT are according to [1] "air pollution, forest fire detection, plant health monitoring, athlete health care, bridges monitoring, control of unauthorized access to prohibited areas, radiation level measurement, traffic optimization, electromagnetic radiation level measurement, smart roads and lighting, city noise maps, smart shopping, leakage of water pipes, diagnosis of motor vehicles, smart parking, waste management, water quality and navigation, and so on."

A connection that exists between IoT and ubiquitous computing should be defined because this term will be used in the following text. Ubiquity computing is according to [2], based on two technologies: IoT and Cloud Computing. It is characterized by the use of very small computers that are built into everyday objects, and which communicate with each other.

2. Basic for IoT

In this paper [3], there has been given a description of the basic technological components needed in order to achieve the interconnection of objects with IoT. The two basic components are IoT Gateway and Device Management (Device Management).

"IoT gateways primarily act as the bridge to connect sensor networks with traditional communication networks, having capabilities such as protocol conversion and device management." [3]. Modern IoT gateways can also partially process the collected data, which means that they don't have to send raw data further but can process data first and send them further afterwards.

Considering that there are many devices that can be plugged into the IoT system, there is a need for selection and later management of such devices. "As a result, the device management frameworks developed by the Open Mobile Alliance (OMA) and the BroadBand Forum (BBF) have been accredited as the standard technology solutions for device management in IoT." [3].

In order for IoT to be successful, technological interoperability, resilience and reliability must be ensured .Communication in IoT system is often wireless. The appropriate protocols are also used. "Specifically, we categorize the protocols into the following groups: application protocols, payload container protocols, messaging protocols and legacy protocols." [3].

3. Application in Education

Some of the applications of IoT in education according to [4] are:

- Distance learning, which allows students to learn from remote locations, and at the same time it allows parents to get information about their children's learning progress.
- Ensuring greater security, for example calling the police, ambulance or firefighters in case of need.
- The operations performed in the educational institution can be automated. These operations are tuition fees, annual reports, student attendance notice, and even the performance of some exams.
- The special needs of some students can be met by using IoT. For example, students who have a problem with vision can get texts with a larger font or audio output.

Ubiquitous computing, and therefore IoT, are able to offer more content and opportunities in the area of education, available at any time and in any place. It is also possible to combine (when learning) real and virtual entities, which will be illustrated below. "The potential of ubiquitous learning is reflected in increasing access to learning content and collaborative learning environments supported by computers anytime, and anywhere. It also allows the right combination of virtual and physical spaces. The purpose of ubiquitous computing technology is basically improving learning processes." [5].

An example of the use of IoT in education is given in [5]. A physical object is connected to one or more virtual objects. A system with three levels is proposed:

- The first level is hardware and it connects physical objects through sensors and similar technologies
- The second level is infrastructure level and it provides possibilities for access to the Internet via 3G or 4G network.
- The third level is the level of applications and services that determine which information will go from physical to virtual object.

The contribution that the mentioned work has, is to improve teaching and learning through IoT. The illustration of the system operation is given in Figure 1. The case study was conducted during one semester, and the course name is Introduction to System Engineering at the University of Cordoba, Colombia. In this course students get familiar with hardware components and their most important operations. Students are first theoretically familiar with computer hardware before any practical work. The behavior of students is tracked. Students receive a text for a piece of hardware on their mobile phone via NFC or QRCODE tag attached to that piece of hardware. When a student is connected to the appropriate hardware, the system sends an animation or video clip to the mobile device that illustrates the basic operations during the operation of that hardware.

The system architecture is shown in Figure 2. On the server side, there are two databases, one containing information about the learning hardware (the teaching material) and the other containing student profiles and their activities during learning. On the client side there is an interface that relates either to HFC or QRCODE. As a response, the teaching material is sent in the appropriate format to the mobile device of the student. If a student performs some activity on this material, the track is sent to the server, and the result of the activities in the student profile is recorded there.

The students were divided into two groups of 25 members, and the results showed that students in the experimental group achieved greater advancement in knowledge, although the knowledge in both groups was equal before the application of this part of the course, which includes IoT.

Fig. 1. An illustration of how the system works [5]

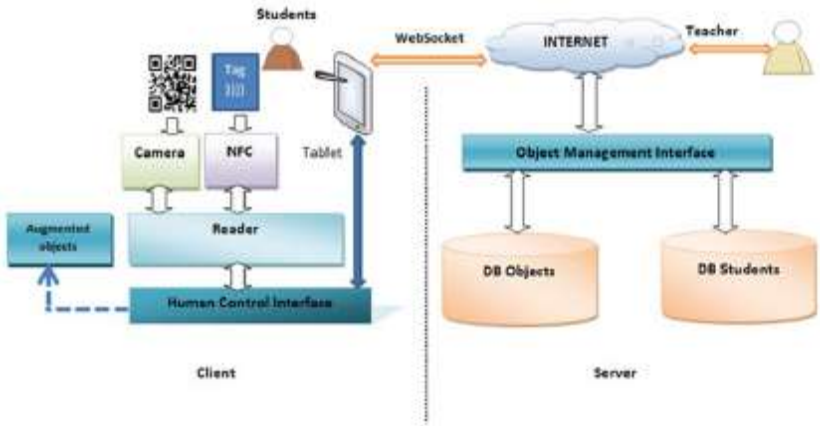


Fig. 2. System Architecture for Internet Objects (Internet Things) [5]

Another example of the use of IoT in education is given in the article [6]. The main motives of this research with Open University in the UK are that new jobs require new

skills, that more and more people in the world want education, and that in the field of technology people are becoming less consumers and more and more manufacturers of technical solutions. The sample included 1967 students and was conducted from October 2011 to March 2012. "The key objective of this new course, called My Digital Life, is to place the Internet of Things at the core of the 1st year computing curriculum and to prime students from the very beginning for the coming changes in society and technology." [6].

The infrastructure used in this course implies

- SenseBoard (Figure 3) which represents a set of necessary devices such as sensors, stepper motors, USB cables ...
- Sense (http://sense.open.ac.uk) is a newly developed visual programming language and development environment (Figure 4) which enables easy operation with SenseBoard.
- Cloud infrastructure allows students to exchange knowledge, experiences, solutions, messages on forums...

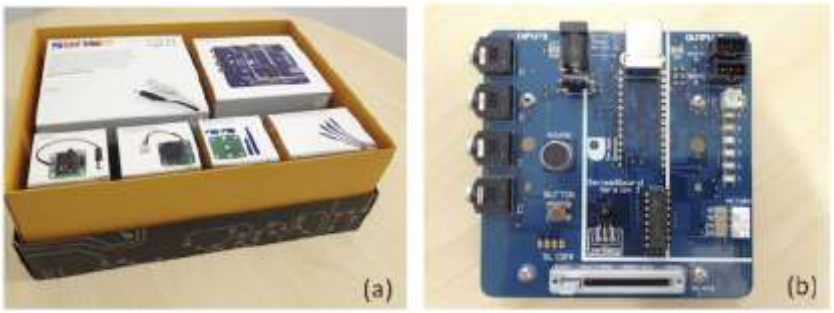


Fig. 3. SenseBoard [6]

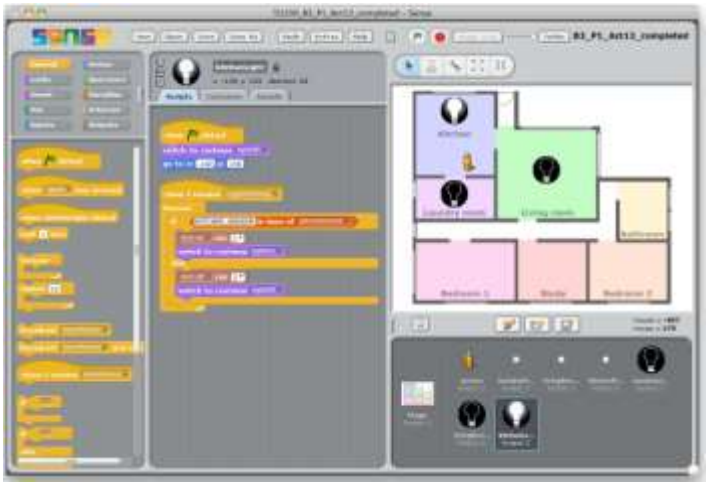


Fig. 4. Sense Programming Environment [6]

After only twenty minutes, some students were able to make their first program. This course has demonstrated a new model of learning in computer education and that it is possible to successfully study the problems of IoT in the first year.

4. Conclusion

IoT is basically a set of devices that are connected by appropriate ways and protocols and together make up the system. In today's world they are very present and their expansion is expected in the future. The application of IoT is in different areas, and one of the areas is also in education. Two examples are given in this work: one that relates to the way IoT helps in the learning process, and the other that gives a successful model to learn IoT skills in the first year of study without some specific knowledge.

References

1. Eremić, Ž., Manojlović L.: Internet of Things (IoT). In Proceedings of the 5th International Conference on Entrepreneurship, Engineering and Management, Zrenjanin, Serbia, 65-71. (2016)
2. Caceres, R., Friday, A.: Ubicomp systems at 20: Progress, opportunities, and challenges. IEEE Pervasive Computing, Vol. 11, No. 1, 14–21.(2012)
3. Gazis, V., et al.: A survey of technologies for the internet of things. Wireless Communications and Mobile Computing Conference (IWCMC), 2015 International, 1090-1095. (2015)
4. Venkat. S.: Internet of Things in Education: The possibilities are numerous (2014) [Online]. Available: http://blog.e-zest.com/internet-of-things-in-education-the-possibilities-are-numerous/ (current September 2017)
5. Gómez, J., Huete, J. F., Hoyos, O., Perez, L., Grigori, D.: Interaction system based on internet of things as support for education. Procedia Computer Science, Vol. 21, 132-139. (2013)
6. Kortuem, G., Bandara, A. K., Smith, N., Richards, M., Petre, M.: Educating the Internet-of-Things generation. Computer, Vol. 46, No. 2, 53-61. (2013)

Solid Waste Collection and Management using Smart Algorithm based on Genetic Programming

Ramona Markoska[1], Goran Mateski[1], and Aleksandar Markoski[1]

[1] Faculty of information and communication technologies, Partizanska bb, 7000 Bitola, Macedonia
ramona.markoska@fikt.edu.mk, goran.mate@yahoo.com, aleksandar.markoski@fikt.edu.mk

Abstract. According actual environmental policies, problem of solid waste collection is one of the significant priorities of many different municipalities. Starting from the map of solid waste containers as an initial condition, the given pilot model in this paper, try to solve the mentioned problem for municipality of Prilep. The main goal of this model is to reduce transport costs, based on control of two parameters: container filling, and routing optimization. For this purpose, depending on the frequency, time and extent of filling, it needs to be created and constantly revised a sub-map with containers, as a subset of primary map. Route optimization for those subset maps, is made using modified smart algorithms from subfields of artificial intelligence.

Keywords: Vehicle Routing problem, optimization, waste management,

1. Introduction

Modern ways of living continuously produces significant amounts of urban solid waste. According Eurostat data Costs for everyday processes of waste collection, and transport to landfill, represent an important financial share in the municipal budget. Each municipality, according ecological and financial points of view, should analyze situation and bring a sustainable multi-criteria waste management process. In order to optimize whole process of collection and transport, routes should be optimized for those subset maps.

The number and location of containers changes periodically, which lead to need of re-calculation of subset of primary map. Continuously, the level of fullness of container is changing, and, this information should be used as a selection criteria, which containers are parts of following route and needs to be emptied. Described activities need to be done for purpose to ensure quality data input for smart algorithms.

All those activities are planned and performed according two main purposes: to ensure quality management of waste collection process from the ecological point of view; and, the same time, to minimize financial implications and costs. For a long time, whole solid waste quantity measurements across the world show a growth trend of quantity,

leading to increased costs of collecting, transporting and disposing. Independently of waste quantity, there is always a possibility to manage with municipal waste collection following some ecological recommendations. Also, there is an opportunity to decrease transport costs, using smart algorithms for optimizing working trajectories of waste vehicles. According Eurostat [1], municipal waste accounts for only about 10 % of total waste generated when compared with the data reported according to the Waste Statistics Regulation. Table 1. present statistical trends, collected and published data on municipal waste since 1995. These data are widely used for comparing municipal waste generation and treatment in different countries, and indicators on municipal waste are used to monitor European waste policies. The data on municipal waste expressed in kilograms per capita are part of a set of indicators compiled annually to monitor the EU's sustainable development strategy, and shown decreasing trend.

Table 1. Municipal waste trends in EU from 1995-2015 [1].

	1995	2000	2005	2010	2015	change (%) 1995-2015
EU-28	:	521	515	504	476	:
EU-27	473	523	517	505	476	1
Belgium	455	471	482	456	418	-8
Bulgaria	694	612	588	554	419	-40
Czech Republic	302	335	289	318	316	5
Denmark	521	664	736	:	789	52
Germany	623	642	565	602	625	0
Estonia	371	453	433	305	359	-3
Ireland	512	599	731	624	:	:
Greece	303	412	442	532	485	60
Spain	505	653	588	510	434	-14
France	475	514	530	533	501	5
Croatia	:	262	336	379	393	:
Italy	454	509	548	547	486	7
Cyprus	595	628	688	689	638	7
Latvia	264	271	320	324	404	53
Lithuania	426	365	387	404	448	5
Luxembourg	587	654	672	679	625	6
Hungary	460	446	461	403	377	-18
Malta	387	533	623	601	624	61
Netherlands	539	598	599	571	523	-3
Austria	437	580	575	562	560	28
Poland	285	320	319	316	286	0
Portugal	352	457	452	516	:	:
Romania	342	355	383	313	247	-28
Slovenia	596	513	494	490	449	-25
Slovakia	295	254	273	319	329	12
Finland	413	502	478	470	500	21
Sweden	386	428	477	439	447	16
United Kingdom	498	577	581	509	485	-3
Iceland	426	462	516	484	583	37
Norway	624	613	426	469	421	-33
Switzerland	600	656	661	708	725	21
Montenegro	:	:	:	:	533	:
The former Yugoslav Republic of Macedonia	:	:	:	351	:	:
Serbia	:	:	:	363	259	:
Turkey	441	465	458	407	400	-9
Bosnia and Herzegovina	:	:	:	332	:	:
Kosovo*	:	:	:	:	178	:

(:) not available

(*) This designation is without prejudice to positions on status, and is in line with UNSCR 1244 and the ICJ Opinion on the Kosovo Declaration of Independence.

Waste collection is the collection and transport of waste to the place of treatment or discharge by municipal services or similar institutions, or by public or private corporations, specialized enterprises or general government. Collection of municipal waste may be selective, that is to say, carried out for a specific type of product, or undifferentiated, in other words, covering all kinds of waste at the same time. [2].
Regarding the use of smart algorithms for vehicle routing optimizing, the last decades, many researchers searching for solutions, have modeled this problem on various ways and versions, such as: vehicle routing problem [3], [4], sometime, for the improvement purpose, combined with other computational techniques, capacitated arc routing problem [5], ant routing problem [6], and different kinds of traveling salesman problem [8]. An smart algorithm modification is presented in this paper, based on symmetric traveling salesman problem [7], which use authentic data for locations of different kinds of waste container [8]. The following sections will explain necessary steps formulating problems and finding solutions.

2. Solid waste collection modeling – Municipality of Prilep case study

According to reports and statistic indicators from the company "Komunalec", Municipality of Prilep, there are a few different kind of containers, depend of garbage and users profiles. For example, there are 20315 households that use small waste bins, and fixed timetable for waste removal (1 time per week, in planned days). But, there also, 33 facilities for educational purposes, such as kindergartens, primary and secondary schools, and colleges, and 1417 different kinds of business entities, which share larger containers, all of those, located according Fig.2.

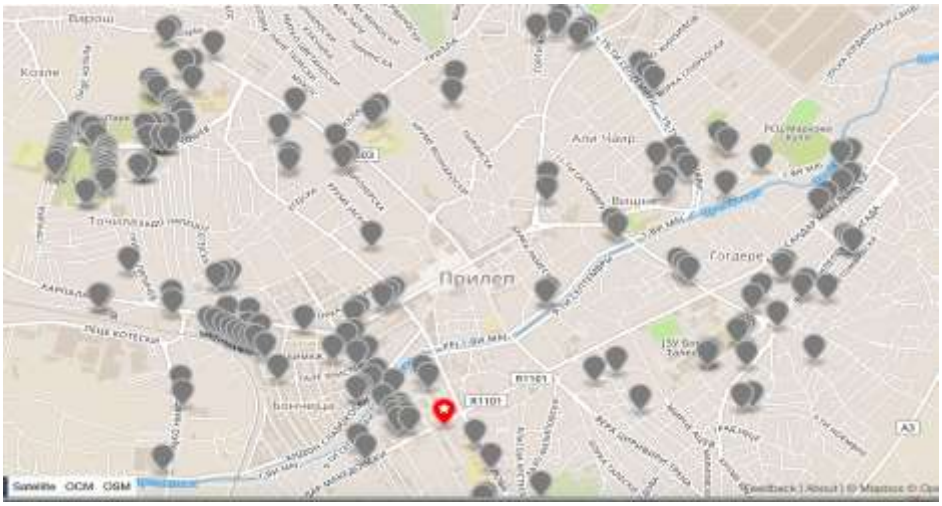


Fig.2. Solid waste containers and bins in Municipality of Prilep

Large and peripheral containers are visited according to a special plan, based on solid waste collection statistic from past years. The containers of central urban area, as can be seen, are naturally grouped by location, and marked on schema. The further assumption in this paper is that the observed streets are two-way, which corresponds to reality. Furthermore, starting from coordinates od map, according to Manhattan metric, which means that it is sufficient to calculate the distance between different groups of containers and use as input parameters on the distance matrix. Any particular container of each group should be equipped with sensor for fullness measurement. Before routing, according to those measurements and capacity of vehicles, it is possible to make simulations and fin optimal number of vehicles which are necessary for collecting the detected waste.

3. Smart Algorithm implementation

The continuous growth of power of processors and computer systems in the last decades, enable a new approach for difficult problems solving. Further, a new chance is given, to solve some mathematical and technical problems, which have high demands for processing, known as "brute-force", on a completely new way. This new possibility, based on biological inspired techniques, is given through paradigm called evolutionary computation.

3.1. Evolutionary computation

All biologically inspired techniques for real problem solving, widely known and described under the term evolutionary computation have some common characteristics. Primary advantage of evolutionary computation is that is conceptually simple, but not always easy for use. That's because every successful implementation needs imagination and ability to translate real model and process on the level to be comparable with natural processes of evolutions. Depending on variations of basic parameters, there are a few basic evolutionary categories, known as: Evolutionary algorithms, Evolutionary strategies, Genetic algorithms and Genetic programming. All of those, have some common functional characteristics, as well:

- Initial population, formed from
- Individuals, represented with chromosomes
- Number of generations planed to convergent to
- Adequate solution,
- Set of operators, like mutation, crossover, tournament selection, used to make changes of initial population,
- Fitness function defined as a convergent criteria, and finally, the
- Problem space which represent union of all possible solutions that satisfy the given conditions.

The problem space is actually made from all individuals which tend to be adequate solution. Each individual, called also, genome or chromosome, have some potential to be solution. The same process of founding solution is made from activities which change genomes and individuals, on way to be best possible adopted according the

fitness function demands. All categories of evolutionary computation, are based on meta-heuristic and stochastic optimization, which means, to seek for optimal solution from problem space. But, unlike the deterministic method, where it is always possible to get a solution with the desired accuracy, stochastic methods do not guarantee finding the global optimum, or the required accuracy. Genetic programming is most advanced technique in evolutionary computation. The main advantage over other evolutionary techniques is the way to generate the chromosome, which represent individuals. Chromosomes created using other evolutionary methods, have only values, but those used in genetic programming, are parsed GP tree on values and functions. The choice of parameters and operators in genetic programming, has goal to make modification of parsed tree, and lead to faster solution.

3.2. Smart algorithm implementation for municipal solid waste collection and routing optimization – Municipality of Prilep

According previous explanations, it is necessary to mark container locations on map of MoP, further used to calculate distance matrix for loading spots, in central area. According to the real situation, some working assumptions are made in this research:

- All used streets in urban area are two-way,
- The containers and bins are equipped with waste level sensors [9]

Problem which has to be solved is: how to make optimal schedule, to visit all loading spots in central area city , depend of following presumptions and needs: to optimize routing process of vehicles, and to find the smallest number of vehicles needed for collecting waste, according the measured level of each particular container. A series of initial experiments has been made, at first, for whole city according Fig.2., and later, according container positions show on Fig.3.

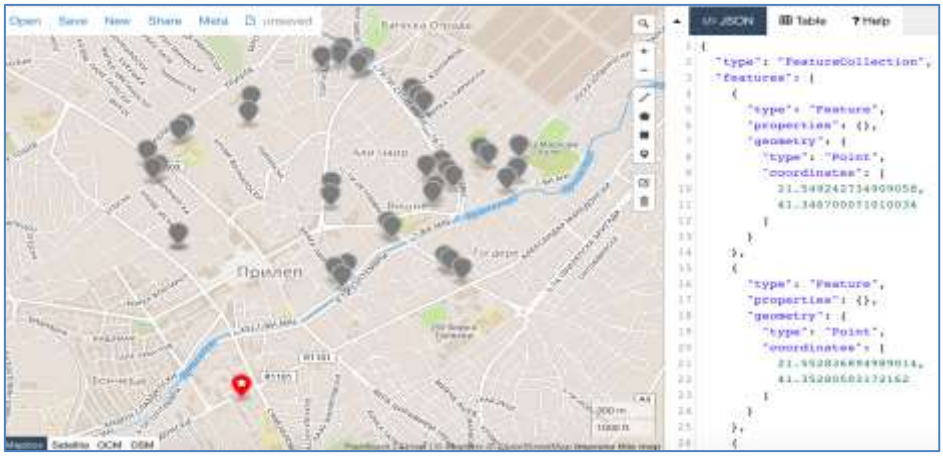


Fig.3. Solid waste containters shema - Gjogdere, Alichair, Visne and part of Center

In situation when a smart VRP algorithm was used, for a whole Prilep town, the following results were obtained: the route is optimized, the total length is reduced, the required number of vehicles depends on the amount of waste. According optimization processes, this leads to a conclusion that the sensors for quantity measurement, are useful. On the other hand, the containers are unevenly distributed, for which, the obtained trajectory, although the shorter are unpopular.

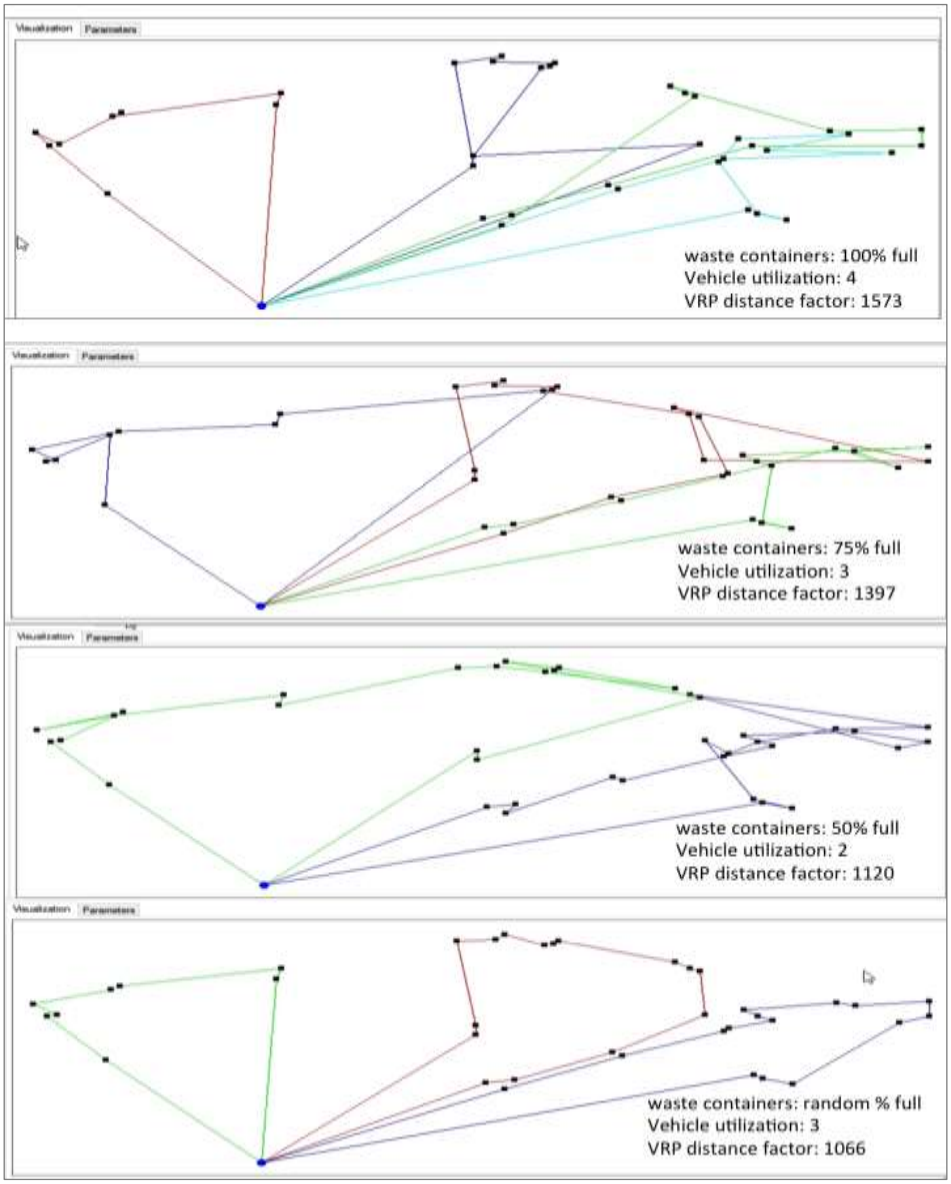


Fig.4. Experimental results – solid waste collecting and routing MoP

Instead of the whole city, it's better to make optimizations for particular quarts, based on grouping waste containers. In this case, it is possible to manage the number of vehicles and optimize routing, and, at the same time, achieve logically acceptable trajectories. In purpose better explain this idea, an quart in Prilep is selected, called AliChair-Visne-Gjogdere, which include 39 container locations. Starting point for routing is location of the company Komunalec, situated in urban area (marked with star on Fig.3.). Open source Leaflet Routing Machine, and GeoJson, are used to pick waste points coordinates (latitude and longitude, Fig.3) and transform it into Manhattan distance. Experimental results of algorithm use are presented of Fig.4.

3.3. Intepretation of results

According to the assumption that the input data on the amount of garbage are available for each container, there are 4 experiment: 100%, 75%, 50% of fill, and random generated amount %, for each container particularly. The data on percentage of container occupancy are assumed. The number of container is continuously 39, according the current situation. Depending of amount of the waste, there are various number of vehicle, 2,3, or 4, which is one more segment of optimization. It is also possible to conclude a continuous shortening of the routes. Furthermore, irrespective of the fact that during the algorithm process, the graphical route views are straight lines, these are calculated according real street geometry in Prilep, using as it said previous, Manhattan distance. Smart algorithms, as mentioned above, give the results that are most appropriately accepted at this point of observation, but not necessarily the best possible. The more factors being taken into analysis, the greater the likelihood and the possibility of optimization.

4. Conclusion

Solid Waste Management is one of the most influential financial processes of each municipal budget. Therefore, any savings in this process are of importance. One of the way how to improve it, is continuous adaptation and improvement of basic processes of waste collection. In the waste management process itself, there are the elements which are variable, which need to be predicted and calculated. Smart algorithms, as explained above, have great potential for use, which is why they are the subject of research for several decades. Presented results in this work has shown advantages and obstacles of smart VRP algorithm use: For better optimization, it is recommended to process container fullness information, city segmentation in the waste collection area, and the route re-calculation before the collection process begins. This work can be upgraded, using real data for sensor measurements, and perform calculation for other city areas.

5. References

1. Municipal waste statistics, Data extracted in July 2017. Most recent data: Further Eurostat information, Main tables and Database, http://ec.europa.eu/eurostat/statistics-explained/index.php/Municipal_waste_statistics#Further_Eurostat_information, last visited 01.09.2017
2. Glossary of Environment Statistics, Studies in Methods, Series F, No. 67, United Nations, New York, 1997.Garcia-Molina, H., Ullman, D. J., Widom, J.: Database Systems: The Complete Book. Prentice Hall, New Jersey, USA. (2002)
3. Laporte, G.:The Vehicle Routing Problem: An overview of exact and approximate algorithms, European Journal of Operational Research 59 (1992) 345-358 345
4. Gendreau.M., Hertz, A., Laporte, G.: A Tabu Search Heuristic for the Vehicle Routing Problem.Management Science, Vol. 40, No. 10 (Oct., 1994), pp. 1276-1290
5. Wohlk, S.: A decade of capacitated arc routing, in The vehicle routing problem – latest advances and new challenges, B. Golden, S. Raghavan and E. Wasil, eds., Springer, N.Y., 2008, pp. 29 - 49.
6. Yu, B., Yang Zhong, Z., Yao B. : An improved ant colony optimization for vehicle routing problem, European Journal of Operational Research 196 (2009) 171–176
7. Narwadi, T. and S.: An application of traveling salesman problem using the improved genetic algorithm on android google maps, AIP Conference Proceedings 1818, 020035 (2017); doi: 10.1063/1.4976899
8. Cirasella, J., Johnson, D.S., McGeoch, L.A., & Zhang, W. (2001). The Asymmetric Traveling Salesman Problem: Algorithms, Instance Generators, and Tests. *ALENEX*.
9. G.Mateski. R.Markoska and A.Markoski, “Development of Low-Cost Wireless Sensor Network for Solid Waste Management”, (submitted for) International conference on Applied Internet and Information Technologies, Zrenjanin, Srbija, October 5 – 6, 2017

Development of a pixel chaser aimbot for FPS games

Predrag Pecev[1], Dejan Lacmanović[1], Dragica Radosav[1] and Branko Markoski[1]

[1] University of Novi Sad, Technical faculty "Mihajlo Pupin", Đure Đakovića bb, 23000 Zrenjanin, Serbia
{pecev, dlacman, radosav}@tfzr.uns.ac.rs
markoni@uns.ac.rs

Abstract. In this paper we examine foundations of building a software for cheating in online games especially in FPS (First Person Shooter) genre, commonly known as aimbots or aim robots. Purpose of an aimbot is to track and target players opponent(s) better than a human player can, in order for a player to score more points. Most of the commonly known aimbots use game data streams and inject certain data in order to achieve aforementioned feature and provide players with "enhanced" aiming skills. Second class of aimbots rely on analysis of what is happening on the players screen and are commonly known as "pixel chasers", since they analyze images (screenshots) that are rendered on players screen. This paper describes development and a basic functionality of a "pixel chaser" aimbot that was developed for research purposes with intention to comprehend how aimbots are made and potentially develop protection methods against players that use aimbots. Furthermore, certain optimizations where implemented in order for aiming to be humanlike as possible without game memory or network data package modification which can be easily traced. This paper also gives common insight on how to detect players who are using aimbots.

Keywords: aimbot, image processing, input simulation

1. Introduction

Topic that this paper covers is somewhat unexplored yet there are papers that give insight into gaming habits and usage of cheating software. Such papers are papers of Su-Yang Yu et al. [1] and Xiang-Bin Shi et al. [2] that give insight on aimbot detection methods.

Work of Chen et al. [5] displays an effort that is made in identifying bots in FPS games by evaluating players movement in game in Quake 2, a game developed by id Software [7]. Using k-NN (Nearest Neighbor) and SVM (Support Vector Machines), with and without dimension reduction, they could differentiate real players movement opposed to a bot movement with rather satisfactory results. Yeung et al. [6] proposed an aimbot detection method that uses a dynamic Bayesian network while Yu et al [4] propose a novel supervised heuristic method based on distribution comparison matrices that can detect players that use aimbots.

However, some countermeasures are often implemented by game developers themselves on the level that do not hinder online game experience. For example, regarding hit detection in FPS games, it can be done on the client side, server side or some combination of aforementioned methods. Since games are played online on remote

servers, network latency is also an issue which can hinder gameplay experience and cause effects such as rubber – banding, poor hit detection etc. In order to provide solid, non cheating experience for most of the players, some games implement data stream validation on the server side which can add more in game latency for the player which can, severely hinder players gameplay experience thus making him stop playing the game in general. Latency in game depends on the location of a game server and network infrastructure between client and server.

In general, it is up to the player do decide on which server will he play based on the previous experiences. However, previous experiences include negative feedbacks regarding the amount of players that use cheating software on game servers and the involvement of server administrators that should filter out players that are either suspected to cheat or confirmed cheaters. This is today mostly done by reporting a player for some suspicious activity and / or providing recordings or demos of played games where a suspicious activity has been seen by other players. Since the main goal of most modern games is to get frag (a gaming term meaning to eliminate the opponent), aiming can be a very difficult task for some players. Aiming in general is influenced by opponents moving patterns, players experience in a game that he or she is playing and players motor skills.

Most of the motor skills for aiming are developed by playing the game, yet some players never develop good aiming practices while some are simply more versatile and talented regarding aiming and object tracking in games. Therefore they simply excel compared towards other players regarding aforementioned feature. In addition to what was previously said, some players with natural disposition and talent towards aiming will most likely always aim better than other players thus having and advantage over them. Since one of the key features of FPS games is the players ability to deal damage to his opponent thus scoring points (frags) when opponents amount of health reaches zero value, aforementioned ability is considered to be very important. Other important factors of a game that is being played are players experience, knowledge of the game, strategies and tactics that he deploys, but in this paper, we put focus on the aiming because the topic of this paper is aiming robots or aimbots that assist players in aiming.

2. Pixel chaser aimbot implementation

Developed aimbot is a pixel chaser aimbot, meaning that the software will “assist the player” or aim towards a certain cluster of pixels that satisfy a certain criteria. Since developed aimbot is classified as a pixel chaser it is clear that the software captures and analyses players screen in constant time intervals and assists player based on the results of a analyzed screenshot.

Developed aimbot is mostly oriented towards high speed competitive FPS games such as Quake 2, Quake 3 Arena, Quake Live, all developed by id Software [7], Quake 4 developed by Raven Software [10], Painkiller developed by People Can Fly [9], Unreal Tournament, Unreal Tournament 2003, developed by Epic Games [8] etc. These games have mostly defined an FPS as a genre, especially early works of id Software [7] such as Wolfenstein 3D and Doom series from which Quake series originated. Most of these games have mods, or add-ons that allow various customizations of user interface without altering game physics, and some of those mods where officially recommended and

accepted as a competitive platform with certain limitations regarding the amount of user interface tweaking that was allowed. For example Quake 3 Arena was played with OSP [11] (Orange Smoothie Productions) mod and later with CPMA [12] mod on large scale official tournaments such as ESWC [14] (Electronic Sports World Cup), QuakeCon [15], CPL [16] (Cyberathlete Professional League) while Quake 4 was played with Q4Max [13] competition mode.

Since in competitive gaming, sole goal of any gamer is to beat an opponent in order to progress further into tournament, quality of graphics are somewhat ignored and downgraded and customizable towards the player as shown in Fig. 1 and Fig. 2. One of the key feature is altering an enemy model so it better suits the player in order for a player to spot him easily. This is achieved by setting a preferred enemy model and color that most of previously mentioned games support. Commonly, a bulky or a large noisy enemy model is picked, and bright colors are applied. In Quake 3 Arena, Keel and TankJr player models are commonly set and dyed bright green as shown in Fig. 1 and Fig. 2. Enemy model customizations in other FPS games will not be shown here since they are mostly based on the same or similar method. Now, this is where pixel chaser aimbots have their appliance.

Fig 1. Bright green colored Keel player model

Fig 2. Bright green colored TankJr player model

Based on the analysis of a captured screenshot, it is possible, to find pixels that match exact certain color, represented by any color model (currently implemented RGB that can be converted to CMYK and HSB), or focus on any color component and find pixels based on a certain complex criteria. Currently aimbot is optimized for duel or 1v1 gameplay, meaning that the aimbot assumes that there will be only one cluster of pixels per frame that can be treated as a target or an opponent. In that case, by default, if a certain number of pixels is found they are treated as an irregular polygon shape and then a centroid is calculated based on the corners of a proclaimed polygon shape. Formulas for centroid point of a polygon are pointed out in formulas marked (1) and (2) where C_x stands for X coordinate of a centroid, C_y stands for Y coordinate of a centroid while A from formula (3) stands for an area of a polygon.

This way, an entire frame, or a screenshot is not fully analyzed since target is proclaimed based on the previously mentioned criteria, providing desired behavior and goal with reduced computational costs.

$$C_x = \frac{1}{6A}\sum_{i=0}^{n-1}(x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \quad (1)$$

$$C_y = \frac{1}{6A}\sum_{i=0}^{n-1}(y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \quad (2)$$

$$A = \frac{1}{2}\sum_{i=0}^{n-1}(x_i y_{i+1} - x_{i+1} y_i) \quad (3)$$

Currently, there is an effort being made towards adapting developed aimbot towards team play game types of previously mentioned games where there can be more than one opponent on a players screen. The idea is that when pixels matching given criteria are found they are to placed in groups based on their proximity. These groups are then to be treated as a an irregular polygon shapes, and then a centroid point will be calculated for each proclaimed polygon shape. From calculated centroid points, the closest one to the players crosshair will be picked as a players target.

Fig 3. Aimbot Control form

Next, having in mind the resolution of a captured screenshot and the fact that in FPS games players crosshair is always located in the center of the screen a crosshair reference point is easily calculated. Coordinates of an aforementioned point represent half of a screenshots width regarding an X axis, and half of a screenshots height regarding a Y axis. In current implementation offsets between X and Y axis values between location of the crosshair and centroid point are calculated and mouse input movements are then

created in order to move players cursor to an enemy's position. This cycle is repeated every 10 milliseconds and enables real-time tracking and locking of enemy models, thus enabling a player to deal larger amounts of damage to their opponent.

Developed aimbot is designed to work with Microsoft Windows platform and it is written using .NET 3.5 platform and based on Windows Application Interface (WinAPI) function calls that are stored in Kernel32.dll and User32.dll files and invoked via Interop Services and DllImport. Screenshot from a players screen can be obtained in two ways:

- By triggering and storing a screenshot in an operating memory (RAM) as a bitmap image
- By triggering and storing a screenshot on a hard drive which requires some additional game configuration

It is clear that capturing storing a screenshot within a computers operating memory is a preferred method since it is much faster and it does not rely on performance of an external memory (hard drive) or some game tampering in order to make previously mentioned screenshot analysis.

Developed aimbot can be configured to be started or stopped on any keyboard key that a player desires, and it works as a simple toggle button. For example, in Fig. 3 Delete (DEL) key starts and stops the aimbot. Also from Fig. 3 it can be seen that, depending on the resolution (Width and Height) in which the player is playing the game, amount of pixels that need to be checked can be quite large, so a player is given an option to provide density of a matrix that is going to be applied when searching for pixels that satisfy certain condition. By entering values (here noted as N) in fields labeled Every pixel by X Axis and Every pixel by Y Axis, every N-th pixel on X and / or Y pixel will be checked against desired condition. By default, aforementioned fields have a value of one so each pixel of a screenshot is checked against desired condition which can be, depending on a screenshot size, a very time consuming process.

Having this in mind, it is advised to take into consideration resolution of a captured screenshot, and depending on the game and players needs to lessen the search matrix density in order to speed up screenshot analysis process. Also there is an option for aimbot to cancel screenshot processing if a certain number of pixels in once cluster is found. On Fig. 3 value for that parameter is labeled Number of Pixels to Find and the feature regarding aforementioned property was described in the earlier paragraphs of this paper and will not be explained here again. Also it is important to emphasize that screenshot analysis begins from top left corner, that is from [0,0], and continues alongside X or Y axis first. By default, pixels along X axis coordinate are checked in sequence, and slowly values for Y axis coordinate are incremented until the end of screenshot is reached. By unchecking Use X Axis First, order of processing pixels is inverted, that is, first pixels alongside Y axis coordinate are checked in sequence and then values for X axis coordinate are incremented until the end of screenshot is reached.

Conditions that pixels in a cluster must satisfy in order to calculate a centroid point are defined in Color properties group from Fig. 3. If Use Exact Color is checked. vales from textboxes A (Alpha) – indicating transparency (which is disabled by default since captured screenshot is not a composite one) , R (Red), G (Green) and B (Blue) are used on the basis od RGB color model to form a hexadecimal representation of an exact color of a pixel that is to be searched for. If aforementioned checkbox is not checked, then pixels that have a value greater that the one entered in specific channel (R, G or B) are

marked and counted for. For example, values in Color properties group in Fig. 3 indicate that pixels that for G channel have a value greater or equal to 175 will be selected, while, values for R and B will be equal to zero. This way, a user can search for pixels that match certain shades of color (in this particular case a certain shade of green) not just the exact ones.

In the previous chapter, in detail, a screenshot analysis is explained with the emphasis on the means of detecting clusters of pixels that represent players opponent thus calculating centroid point towards a point should be aiming. As previously said, when a point where a player should be aiming for is determined, offsets along X and Y axis are calculated and then trough WinAPI INPUT structure, and SendInput function input is generated and sent to the operating system thus moving users mouse cursor and aiming towards determined point.

This can be done instantly, thus making generated mouse movements look robotic since there would be a instant straight line movement from point A to point B, or optimized using randomized Beziers curves. This feature can be activated by simply checking a Use Bezier Curve checkbox. By default, aforementioned checkbox is not checked, so default behavior of the aimbot is to look and feel slightly robotic.

When using randomized Beziers curves, a series of points based on randomized Bezier curves are formed, and over a period of 1000 milliseconds sent as input thus simulating more human-like movement. However during the tests it was found that calculating Bezier curve points heightens computational costs of operations that need to be executed and the provided result, regarding mouse movement still looked robotic. Based on what was previously said it is clear that in this particular case Bezier curves did not provide solution that would make generated mouse movement more human-like, so as one of the further research goals, finding an appropriate algorithm for replicating human-like mouse movement was set.

However, developed software has several issues that are caused by graphics driver issues and the API that the game itself uses in order to draw itself on the users screen. For some games and versions of Windows operating system, screenshots cannot be made in full screen mode, so a windowed mode was required. Described limitations are somewhat justified since we did not use any DirectX or OpenGL API calls to inject some code and capture screenshots when games are running in windowed or full screen mode.

3. Analysis, testing and performance

Controlled testing of developed aimbot was conducted using a Quake 3 Arena, a game that was developed by id Software [7] in a facilities of a “Magus” internet cafe that was located in Zrenjanin, Republic of Serbia. There where two kinds of players: players that have played aforementioned game before, and players that have never played aforementioned game before or have played it only a couple of times, so their experience with the game was quite limited.

During controlled testing of a developed aimbot players could tell in most cases that they are playing against an opponent with an aimbot due to two reasons. First one is that the players mostly knew one another, their gameplay styles and aiming capabilities, and the second one is that rotation of the enemy model in game was robotic and snappy so

sharp movements where displayed on enemy player models that rose suspicion. Also in spectator mode robotic movement can be easily seen and reported. However players that where not that familiar with the game in most cases did not complain or report anything suspicions since they thought the behavior they experienced was characteristic to the game they played.

4. Conclusion

Developed software made an effort towards understanding how cheating software works while building one. A clear effort was made in designing developed software as an application that does not tamper with game memory or its data while its running. Also a small test was conducted to determine performance of a developed aimbot and the results, based on the oral survey pointed out that it was easily detectable.

Furthermore, many games on consoles such as Microsoft XBOX or Sony PlayStation have an auto snap feature in FPS games which assists the player in aiming with while using a controller by snapping to the enemy that is closest to the players crosshair. This feature can be enabled or disabled. Since consoles where not initially intended for FPS games, controller seems to be quite inefficient regarding aforementioned genre of games. However, by adding analog hat sticks and other optimizations such as auto snapping some improvements where made and FPS genre became quite playable on consoles when a player gets a hold of controller mechanics.

Developed software is classified as a cheating software and highly prohibited if a player is competing on a professional level. Authors of this paper do not encourage usage of cheating software. However, with slight alterations, developed aimbot could be, to some extent implemented in order to assist people with disabilities to play games. In paper Pecev et al. [3] an embedded Arduino based assistive device was described which enables people with disabilities to interact with a computer. By combining these two efforts a simple software that provides ease of access could be made that would provide greater and more dynamical emersion of people with disabilities into some of the famous and highly ranked video game world. These worlds do not have to be classified as FPS games, yet there is an expansion of MMORPG (Massive Multiplayer Online Role Playing Game) and MOBA (Multiplayer Online Battle Arena) games for which, developed “pixel chaser” aimbot could be modified in order to provide people with disabilities more efficient means of interacting with others in virtual worlds. Aforementioned goal is also set as one of the further research goals.

References

1. Su-Yang Yu, Nils Hammerla, Jeff Yan, Peter Andras “A statistical aimbot detection method for online FPS games”, The 2012 International Joint Conference on Neural Networks (IJCNN), 10-15 June 2012, DOI: 10.1109/IJCNN.2012.6252489 , Brisbane, QLD, Australia
2. Xiang-Bin Shi, Xiao-Xue Zhou, Fang Liu, Ying Wang, “A Cheat-Detection Method Based on Fuzzy Synthesis Decision for Aimbot Cheating”, 2010 International Conference on Internet

Technology and Applications, 20-22 Aug. 2010, DOI: 10.1109/ITAPP.2010.5566408, Wuhan, China
3. Predrag Pecev, Dejan Lacmanović, Zdravko Ivanković, Branko Markoski, "Implementation of embedded assistive device for disabled people" International Conference on Applied Internet and Information Technologies, 2016, Bitola, Macedonia, pp 76 – 81, DOI:10.20544/AIIT2016.10
4. Yu SY., Hammerla N., Yan J., Andras P. (2012) "Aimbot Detection in Online FPS Games Using a Heuristic Method Based on Distribution Comparison Matrix" Neural Information Processing. ICONIP 2012. Lecture Notes in Computer Science, vol 7667. Springer, Berlin, Heidelberg, pp 654-661 DOI: 10.1007/978-3-642-34500-5_77
5. Chen, K.T., Pao, H.K.K., Chang, H.C.: "Game bot identification based on manifold learning".. Proceedings of the 7th ACM SIGCOMM Workshop on Network and System Support for Games, pp. 21–26 (2008)
6. Yeung, S.F., Lui, J.C.S., Liu, J., Yan, J.: "Detecting cheaters for multiplayer games: theory, design and implementation", Consumer Communications and Networking Conference, pp. 1178–1182 (2006)
7. id Software, Inc. http://www.idsoftware.com/
8. Epic Games, Inc. http://www.epicgames.com/
9. People Can Fly. http://peoplecanfly.com/
10. Raven Software. http://www.ravensoftware.com/
11. Orange Smoothe Productions. http://www.orangesmoothie.org/
12. Challenge ProMode Arena – CPMA - http://playmorepromode.org/
13. Q4Max Mod. http://www.moddb.com/mods/q4max
14. ESWC http://www.eswc.com/
15. QuakeCon http://www.quakecon.org/
16. CPL – Cyberathlete professional league - http://thecpl.com/

Management and Measurement of Performance in Research Institutes

Goran Petrović [1], Laslo Šereš [2]

[1]Institute of Field and Vegetable Crops Novi Sad, Maksima Gorkog 30
21000 Novi Sad, Serbia
goran.petrovic@nsseme.com
[2]Faculty of Economics Subotica
Segedinski put 9-11, 24000 Subotica, Serbia
laci@ef.uns.ac.rs

Abstract. The main aim of research institutes is the development of both the fundamental and applied science. Although state authorities recognize the common interest in establishment of research institutes, financial support is only partly provided by their founders. This trend is particularly noticeable within agricultural research institutes, which valorize their research results i.e. new cultivars by offering them on the market. The fact that research institutes generate most of their profit by selling their goods and services on the market poses the need to redefine their main objectives and build adequate management models which would reflect duplicity of their existence as both profit and nonprofit organizations. This paper aims to examine the effect that such a status might have on management models in research institutes, focusing especially on the need to build performance management models and identify specific indicators of success as institute performance indicators.

Keywords: institute, management, performance, work process

1. Introduction

Research institutes are established as specialized scientific and educational institutions dealing with specific research areas. Besides the educational purpose, their task is to produce significant new research findings, disseminate research results and transfer them into practice. Generally, research institutes receive their income from institutional or project financing, patent rights or from some other combination of models. However in practice, the abovementioned models of financing can be insufficient or even non-existent, thereby forcing research institutes to adopt the concept of applied science and market their goods or services. The model of organization in research institutes is automatically affected by the adoption of the corporate business model with a different set of problems, priorities and activities. The question is whether research institutes and other research centers can, in that specific case, be satisfied by measuring performance characteristic of research institutions, or they require new, customized performance measurement, including the main purpose of their establishment on one hand, and market / business objectives on the other.

2. Performance

2.1 Performance Management

'If You Can't Measure It, You Can't Manage It'. In short, performance management requires knowledge about:

- our current position and our real possibilities,
- our goal, and how to achieve our goals,
- how to measure results (comparing targets vs. actual values),
- what to do if results are below the expected level (corrective measures).

Defining aims and monitoring their implementation is a must. Performance can be defined as the success of an organization in achieving its set goals [1]. Performance management of an organization is the foundation of business success, based on the definition and monitoring of a number of *key performance indicators* (KPI). Achievement of goals is monitored by performance measurement and comparison with the defined aims and historic data stored at a company database. Corrective measures cannot be defined without measurements. Contemporary research focuses on determining a set of indicators as a tool for measurement of process performance i.e. organization overall performance. Regarding the above, elements of performance management system would be:

1. Determination of stakeholder requirements
2. Establishment of aims
3. Determination of key success factors (KSF)
4. Definition of performance that needs to be measured and measurement units
5. Definition of measurement point
6. Establishment of Roles and Responsibilities Matrix
7. Productivity measurement
8. Analytical processing and reporting
9. Productivity estimates
10. Corrective measures

2.2 Performance Measurement

The choice of performance measurement model depends on the needs of specific measurements in a particular organization. A set of different methodologies can be found in literature sources and listed here are the ones that are most often cited [2]:

- Data envelopment analysis – DEA (Charnes, Cooper, & Rhodes, 1978)
- Theory of Constraints – TOC (Goldratt, 1984)
- Performance Measurement Matrix – PMM (Keegan, Eiler, & Jones, 1989)
- Management Dashboard (De Guerny, Guiriec, & Lavergne, 1990)
- Performance Pyramid (Judson, 1990)

- Activity Based Costing – ABC (Cooper, & Kaplan, 1991)
- Results–determinants framework (Fitzgerald, Johnston, Brignall, Silvestro, & Voss, 1991)
- EFQM model proposed by European Foundation for Quality Management
- DOE/NV model proposed by U.S. Department of Energy Nevada Operations Office (Bellman, Droemer, Lohmann, & Miller, 1994)
- TQM model (Sinclair, & Zairi, 1995)
- Balanced Scorecard – BSC (Kaplan, & Norton, 1996)
- Brown model (Brown, 1996)
- SCOR model created by Council of Supply Chain Management Professionals (CSCMP), 1997.
- Demand to Measure model – DtM (Ljungberg, 2002)
- Kanji's Business Excellence Model (KBEM) (Kanji, 2002)
- Business System Design Decomposition – BSDD (Taticchi, Cagnazzo, Santantonio, & Tonelli, 2010)

The most widely used and mentioned model in literature and practice is *Balanced Scorecard* (BSC). Kaplan & Norton methodology analyzes the management of an organization's performance through four perspectives: Financial perspective, Perspective of the consumer / user, Perspective of internal processes, Perspective of learning and development.

BSC is a system management tool which organizations use for determining the desired goals, maintaining compliance of everyday work with strategic objectives, prioritizing products, services and projects, i.e. monitoring progress and measure performance in accordance with the set aims. It is not a tool for defining the strategy, but for the transfer of strategy into action [3]. It is important to point out that the measurement of performance processes in an organization is not just a collection of the predefined sets of data and their comparison with the defined aims, but that it relies on a comprehensive system management and holistic approach to the use of techniques and tools in order to adjust a product or service to the requirements of stakeholders i.e. people or groups who have an interest in the performance or success of an organization[4]. In addition to the above mentioned methodologies, organizations can use different *Business Process Modeling* (BPM) and *Business Intelligence* (BI) software tools to support the modeling process, data collection and processing, i.e. reporting and performance measurement.

2.2.1 Performance Measurement for Corporate Systems

Performance measurement is a set of activities associated with a goal. Different industries have specific aims derived from the diversity of business specific processes. The specific aims require determination of the appropriate branch metrics. For example, National Bank of Serbia has defined the criteria for measuring the performance of insurance companies and banks according to the international methodology of the IMF called CARMEL (C: Capital adequacy, A: Asset quality, R: Reinsurance and actuarial issues, M: Management

soundness, E: Earnings and profitability, L: Liquidity). These 6 quantifiers express the efficiency of insurance company activities, basically including the analysis of liquidity, efficiency and profitability. Insurance companies need to dedicate special attention to financial statements as a basic tool for determining their business efficiency, avoid insolvency and preserve their market position. This set of metrics is practically the same for banks, with one difference: banks use indicator S (Sensitivity - sensitivity to market risks) instead indicator R, so the methodology is referred to as CAMELS [6].

Branch metrics in *oil industry* focuses on the length of transport routes, estimation of mineral resource deposits, oil stocks, production indexes, pipeline maintenance costs, as well as other parameters generally inherent in companies.

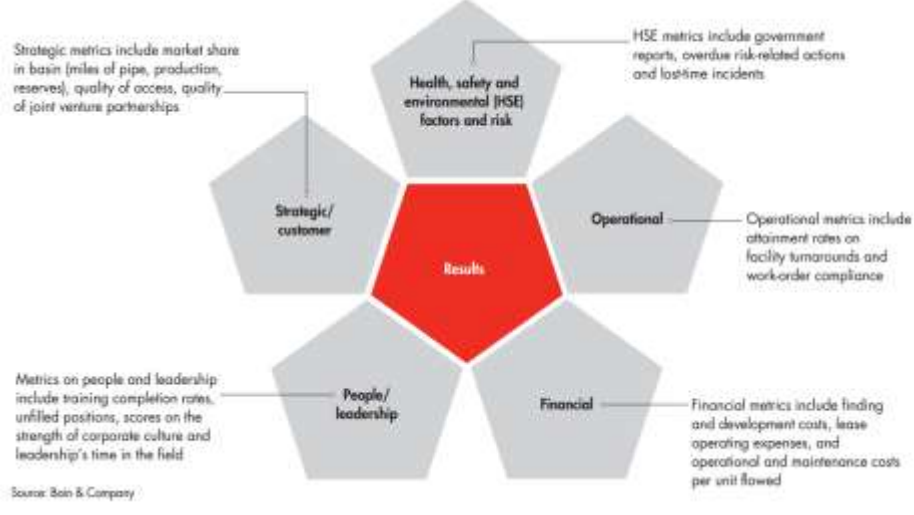


Fig. 1 KPI values in oil industry [5]

Companies whose main activity is agriculture and seed production, view as their key production indicators (KPI) the total yield per season, average yield per cultivar, seed yield and seed production of each seed producer, average price per ton per cultivar, indicators of seed quality (mean germination, waste percentage, etc). Key production indicators in departments of processing and storage are man-hours per ton of processed seed, price per ton of processed seed, energy costs per ton of processed seed, profit/loss per cultivar, stop time of equipment, losses due to stock adjustment, and so on.

KPI process in other industries can also be defined by the same logic. Every business-oriented organization has its own specific KPI indicators and measures the values characteristic of maintaining high quality product and service levels. At the same time, all organizations which offer their goods and services on the market have a common set of

indicators related to performance measurement of product placement, marketing, financial indicators, market share, customer relations (especially social engineering in recent times), which collectively provide management input parameters.

2.2.2 Performance Management in Research Institutes

Accredited research organizations in the Republic of Serbia can be divided into three groups [8]: faculties and universities, institutes, centers of excellence.

The term scientific research refers to the systematic research work carried out with the purpose of obtaining new knowledge, thereby raising the level of general knowledge by using this knowledge in all areas of social development [9]. These organizations are not primarily established as profit institutions, but as research and development institutions with an educational purpose. Revenue is provided by institutional or project financing, although introduction of the combined model has been announced a few times so far.

Indicators used in institutions of high education can be classified according to activities [10]: quality of lectures, quality of research, quality of management activities.
Indicators specific for faculties and universities focus on measuring the value of the main activity, such as student number and satisfaction level, evaluation of the teaching process, qualifications of teaching staff with permanent employment, student / teacher ratio, students' age and type of studies, teaching / non-teaching staff relationship, accreditation standards, and so on. Special attention is given to the teaching-learning process performance indicators which can further be divided into: surrounding area, budget allocated for research and education, institution resources, and achieved results.

It is possible to draw an analogy between faculties and institutes which place emphasis primarily on the educational process, and then the research. Indicators of research work carried out at research institutes could be classified as:

- *Research paper production indicators*
    - number of published research papers according to the number and structure of research staff, the quality of published research papers in terms of publications in which they were published, the number of cited research papers

- *Project realization indicators*
    - number of research projects and their structure, ratio of realized vs. unrealized projects, the number of engaged associates

- *Indicators of applied research within main activity*

For example, institutes have the following possible seed production indicators:

- number of cultivars in preliminary, intermediate and advanced testing
- number of cultivars undergoing registration
- number of released cultivars per year
- total number of released and registered cultivars

The reality of research work in Serbian research institutes is such that it is hardly possible to sustain research activity by relying on projects and science. Research institutes are thereby forced to make profit from market sales of goods and services. The combined model of functioning (both as scientific research institutions and as companies) where institutes compete on the market, introduces additional metrics with the following indicators:

- *Market trade indicators*

The key words here are market and competition. Competition affects the behavior of an organization as a whole, and introduces metrics and key performance indicators of market trade as management decision-making tools. Performance measurement becomes increasingly important when the percentage of institute funding is greater than market trade share. In that regard, there is little difference between institutes and companies. In a situation where the *primary source* of financing is sales of goods and services, introducing market trade indicators, is no longer the fundamental issue. What becomes important is the manner in which an institute operates, due to danger from having a predominantly commercial purpose rather than scientific development. This is especially important in state-owned institutes because it changes their essence and the purpose of their establishment.

3. Examples of Combined Models of Measuring Performance

Performance measurements of the Brazilian Center for Agricultural Research EMBRAPA show that a set of indicators can be designed depending on the requirements. EMBRAPA has 46 research centers in its structure, with one headquarters [11]. Measurements were carried out to determine the index of performance for each individual center, while three input and up to 30 output indicators were defined for this purpose. The purpose of activities is to monitor and control the production of key output variables, such as the production of research papers and transfer of technology into the agricultural sector. [12].

Input KPI indicators are:

- Staff costs (salaries, per diems, fees)
- Operative costs (material, services, travels)
- Capital (amortization-measured)

Output KPI indicators are classified as:

a) Production of research publications
b) Production of technical papers
c) Development of technologies, products and processes
d) Diffusion of technology and image

Research institute Dr. Josif Pancic, which specializes in medicinal plant research, serves as an example to show that research work, research topics and project tasks can take a number of directions within the same research institution [13].

Research in the *pharmaceutical* field are:
- Pharmacognostic studies of medicinal plants (LAB);
- Studying the resources of wild LAB;
- Ethnobotany and ethnomedicine;

Research within the field of *agronomy* are:
- Improving technology of growing LAB in agricultural production;
- Development of technology of organic LAB cultivation;
- Introduction and domestication of protected and endangered LAB species;
- Improvement and development of seed production and LAB breeding;

Another agricultural institute engaged in applied research of *seed production* is Institute of Field and Vegetable Crops Novi Sad [14]. It consists of departments of maize, soybean, small grains, forage crops, industrial crops (oil crops, sugar beet), and alternative plants, including the experimental station in Srbobran, as well as processing and storage facilities. Besides participating in market competition, the Institute offers the services of two laboratories: Laboratory for Soil and Agroecology and Laboratory for Seed Testing. The departments are characterized by different work processes, operating as relatively independent entities with their own specific characteristics due to the diversity of cultivated crops. Clearly, the definition of specific metrics typical of the crop is required.

It is undisputable that institutes operating in various fields of research have different metrics, and the abovementioned examples confirm that there exists no single perspective on performance measurement, not even in the same sector (area of research) at one Institute. The set of indicators will be one in an institute founded primarily (or exclusively) as a research institution, focused on scientific contribution and funded from the state budget, whereas indicators would have to include elements of the corporate governance if an institute focuses on applied research which results in sales of products and services. In both cases, a holistic approach and a thorough study of an organization is required in order to define objectives and create plans and models of performance measurement process, according to dominant activities and the organization as a whole.

4. Conclusion

Research and development centers, institutes and higher education institutions are not initially profiled as for-profit organizations with the primary goal of acquisition and increase of assets. However, global trends, migration of capital, exchange and availability of information, easy access to affordable services and strong competition in all areas have a significant impact on business in these organizations. The global economic crisis and the governance of Serbia through transition affects the funding, reduces it, and leads institutes to the brink of survival. Large state-owned institutes essentially do not differ from state-

owned enterprises. Both are large systems with entrenched habits, slow in decision-making and most often unprepared for market changes and situations which require a fast response to the altered conditions. It is therefore necessary to develop a mechanism and introduce a set of tools for monitoring, implementation, and correction of the current performance, as well as a proactive approach in order to improve performance. Definition of *process maps*, identification of *critical success factors*, determination of measurable *key performance indicators*, the use of Business Intelligence software tools, are all necessary steps and tools typical of corporate governance, applicable to any system that has an organizational structure and defined goals. Activities and efforts are aimed at establishing an orderly system and measurable performance of the whole organization, whereas outputs help the management in making strategic decisions.

Research institutes that are mainly financed through market production and sales of goods and services inevitably require metrics of corporate governance. The presence of competition relocates them from the safety zone of scientific research. Research institutions which are engaged in a dynamic market without having passed some kind of transformation (ownership, organizational, personnel, software), and those which maintain a rigid attitude concerning their primary profession by not taking into account the elements of business strategy, will most likely cease to exist. Only the flexible ones, able to follow trends using all available tools and to make timely alterations, have a chance of survival.

References

1. Balaban N., Ristić Ž., (2013).Upravljanje performansom
2. Simeunović B., Slović D., Radaković Andreja J., (2015). Analiza modela za merenje performansi procesa
3. Medaković Đ., (2010). Značaj Balanced scorecarda za praćenje uspešnosti poslovanja u malim i srednjim preduzećima
4. Golubović D., Radović M., (2011). Postupak merenja performansi procesa
5. Bertocco R., McCreery J.(2014).Operational excellence: Managing performance in the oil and gas industry, Bain & Company, Inc.
6. Jovetić S., (2015). Merenje performansi preduzeća
7. MacRobert J.F, (2009). Seed Business Management in Africa. Harare, Zimbabwe, CIMMYT
8. Ministarstvo prosvete, nauke i tehnološkograzvoja,(avg 2017).http://www.mpn.gov.rs/nauka/nio/
9. Zakon o naučnoistraživačkoj delatnosti ("Sl. glasnik RS", br. 110/2005, 50/2006 - ispr., 18/2010 i 112/2015)
10. Tatjana D. Janovac, (2014). Unapređenje kvaliteta visokoškolske ustanove na osnovu kriterijuma potreba korisnika
11. Brazilian Agricultural Research Corporation - EMBRAPA (avg 2017). https://en.wikipedia.org/wiki/Brazilian_Agricultural_Research_Corporation
12.Geraldo da Silva e Souzaa, Eliane Gonçalves Gomes,(2015). A Performance Measure to Support Decision-Making in Agricultural Research Centers in Brazil
13. Institut za proučavanje lekovitog bilja Dr Josif Pančić, (avg 2017), https://www.mocbilja.rs/
14. Institut za ratarstvo i povrtarstvo, (avg 2017), http://nsseme.com/

On the Concepts of Data Virtualization: NoSQL Systems and Machine Learning Techniques for Performance Modeling

Mimoza Mijoska[1] and Filip Mitrevski[2]

[1] Faculty of Information and Communication Technologies, Partizanska bb, 7000 Bitola, Macedonia
mimozamijoska@yahoo.com

[2] Faculty of Computer Science and Engineering, Rugjer Boshkovikj 16, P.O. Box 393, 1000 Skopje, Macedonia
fmitrevski@gmail.com

Abstract. "Data lakes" has become the new buzzword for organizations dealing with never-ending, increasingly complex and costly requirements for using all of their data sources to produce value. Big Data (new formats of frequently unstructured sensor and click-stream data) needs to be integrated with small data (traditional structured data from transactional or analytical systems). A class of novel data storage products, able to cope with Big Data, are subsumed under the term NoSQL databases, many of which offer horizontal scalability and higher availability than traditional relational databases, or other useful properties, by sacrificing querying options and consistency guarantees. Performance modeling approach aims to predict the performance of a NoSQL database by constructing a number of machine learning models. This review paper presents the concepts of NoSQL systems. Also, it engages in machine learning techniques for performance modeling for NoSQL databases. The literature review shows that a number of experiments have been carried out in order to evaluate this approach, where the target performance metric is response time per second of the request successfully executed by the system.

Keywords: NoSQL systems, data virtualization, machine learning, big data.

1. Introduction

In recent years, the amount of useful data in some application areas has become so vast, that it simply cannot be stored or processed by traditional database solutions [5]. However, relational vendors modified relational systems with object-oriented features and extended SQL to capture most of the advantages of object-oriented systems, while preserving the high performance and strengths of the relational model and SQL. Security and access issues further muddy the waters. Attempting to normalize and integrate data streams and systems in advance of application development, to suit a limited number of use cases is rather impractical and costly. At the same time, data

quality and data supply chains need to be managed, which requires pre-processing to reveal meaningful patterns. External data sources vary in quality and require cleansing before integration with trusted and verified internal sources. Bringing these data sources together to produce insights and value, requires a combination of technology infrastructure, enterprise data modeling, and governance processes [1]. A class of novel data storage products able to cope with Big Data are subsumed under the term NoSQL databases, many of which offer horizontal scalability and higher availability than traditional relational databases or other useful properties, mainly by sacrificing querying options and consistency guarantees [5]. Data virtualization is a mechanism for realizing the data quality initiatives and data governance programs. With data virtualization, master data can be pulled from the enterprise source-of-truth and can be combined with operational and real-time data, to provide a unified view of the target object or process. Virtualization allows for separating the semantic layer (terminology, translations, data definitions, etc.) from multiple data sources [1]. Forrester defines data virtualization as the integration of any data in real time or near-real time from a disparate structured, unstructured, and semi-structured data sources, whether on-premises or cloud [2], into coherent data services that support business transactions, analytics, predictive analytics, and other workloads and patterns.

2. Categorizing the Types of NoSQL

NoSQL systems have been proposed to tackle applications and problem domains poorly served by relational databases. These domains are primarily Big Data domains involving web data such as supporting millions of interactive users or performing analytics on terabytes of data, such as web logs and click streams. The data in these domains is semi-structured, variable and massive. Creating a suitable relational model may prove to be difficult [7].

NoSQL databases draw upon several concepts and techniques to realize flexible data modeling, associated query languages and horizontal scalability. These concepts include, but are not limited to shared-nothing architecture, Hash trees, consistent hashing, REST API, Protocol buffers (Protobuf), Apache Thrift, JSON and BSON, BASE, MapReduce, vector clocks, column family, keyspace, memory-mapped files, and the CAP theorem [3].

2.1. The CAP Theorem

Consistency, availability and partition tolerance (CAP) are the three primary concerns that determine which data management system is suitable for a given application. The consistency feature guarantees that all clients of a data management system have the same view of data. Availability assures that all clients can always read and write. Finally, partition tolerance ensures that the system works well with data that is distributed across physical network partitions. The CAP theorem states that it is impossible for any data management system to achieve all these three features at the same time [3] (Fig. 1).

Fig. 1. CAP Theorem

2.2. Key-Value Databases

The defining characteristics of key-value databases include real-time processing of Big Data, horizontal scalability across nodes in a cluster or data centers, reliability and high availability. Their use cases include applications where response is needed in milliseconds. They are used for session management for Web applications; configuration management; distributed locks; messaging; personalization of user experience and providing engaging user interactions in social media, mobile platforms and Internet gaming; real-time bidding and online trading; Ad servers; recommendation engines; multi-channel retailing and e-Commerce.

Representative systems in this category include Memcached, Aerospike, Redis, Riak, Kyoto Cabinet, Membase, Amazon DynamoDB, CouchDB, BerkeleyDB, EHCache, Apache Cassandra, and Voldermot [3].

For example, the replacement of an RDBMS with Redis for user profile management (100K+ profiles with over 10 filter attributes) was reduced from 3 - 6 sec to 50 ms [8].

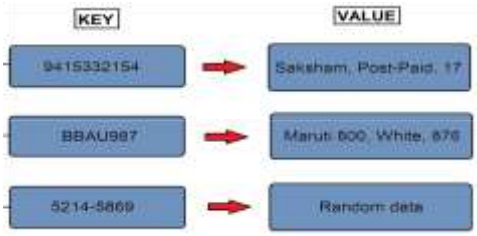


Fig.2. Key-Value Storage System

2.3. Table-type/Column Databases

A columnar database is like an RDBMS with an index on every column, without incurring the overhead associated with the latter. It is also useful to think of column databases as nested key-value systems. Recall that a column family is a set of related columns. Column databases require pre-defining column families, but not columns. A column family may contain any number of columns of any type of data, as long as the latter can be persisted as byte arrays. Columns in a family are logically related to each other and are physically stored together. Performance gain is achieved by grouping columns with similar access characteristics into the same family. Systems in this category include Google BigTable (available through Google App Engine), Apache Cassandra, Apache HBase, Hypertable, Cloudata, Oracle RDBMS Columnar Expression, Microsoft SQL Server 2012 Enterprise Edition [3]. Fig. 3 illustrates how data is stored in a column family database [6].

Fig 3. Column Family Database

2.4. Graph Databases

A graph data model is at the heart of graph databases. In some applications, relationships between objects is even more important than the objects themselves. Twitter, Facebook, Google, and LinkedIn data are naturally modeled using graphs. In addition, graph data models are used in other industries including airlines, freight companies, healthcare, retail, gaming, oil and gas. Graph databases are also popular for implementing access control and authorization subsystems for applications that serve millions of end users.

Graph databases include FlockDB, InfiniteGraph, Titan, Microsoft Trinity, HyperGraphDB, AllegroGraph, Affinity, OrientDB, DEX, Facebook Open Graph, Google Knowledge Graph, and Neo4J [3]. Location based services use this database type, to find mutual friends on Facebook or to establish the shortest paths through everyday traffic (Fig. 4). Examples of Graph Databases include InfoGrid and Neo4J [4].

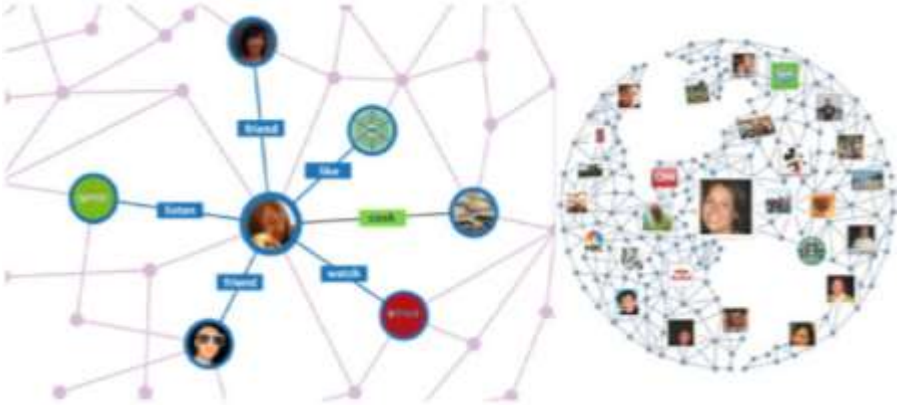


Fig. 4. Graph Database

2.5. Document Databases

These databases are not the document/full-text databases in the traditional sense. They are also not content management systems. Document databases are used to manage semi-structured data mostly in the form of key-value pairs packaged as JSON documents [3].

Document Databases use entire documents of structured data files, such as XML or JSON as datasets. Each record and its associated data are usually stored together in a single document; this simplifies data access and reduces the need for joins or complex transactions. Documents stored are schema-free and similar to each other – this flexibility can be particularly helpful for modeling unstructured data [4].

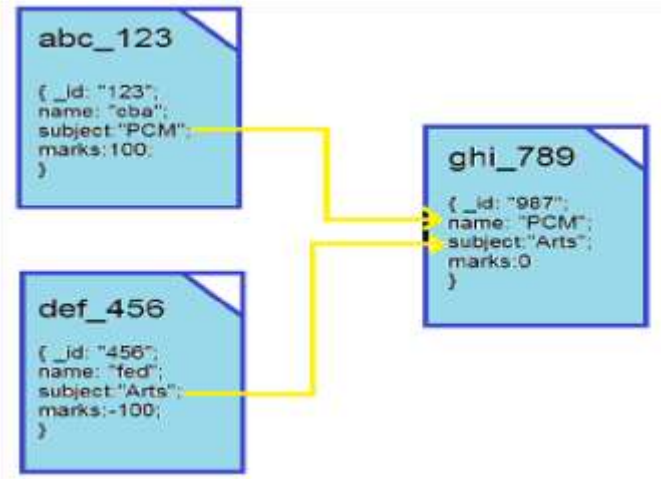


Fig. 5. Document Store Database

3. On the Machine Learning Techniques for Performance Modeling

Various performance modeling approaches aim to predict the performance of a NoSQL database, by constructing a number of machine learning models. For example, under the assumption that *performance predictability* is "the ability to estimate the impact of concurrent query execution on the performance of individual queries", Farias *et al.* [10] have designed experiments considering the following goals:

- Evaluate the predictive power of machine learning models with a variety of regression methods;
- Determine the tradeoff between the adoption of linear or quadratic features.

3.1. Example Experimental Setup

For the experiment approach, a prototype of using Python and Bash Script languages was implemented [10]. The machine learning methods implementation were provided by *scikit-learn* [11]. The benchmark was set with 40 client threads, 1,000,000 records on database, 1-byte size record, and uniform access pattern.

The test database in the experiments is MongoDB [12] set on Master/Slave replication mode. The *ReadPreference* parameter is set to *Nearest*, which means that MongoDB driver sends read for a randomly chosen node, while writes are always issued to the master. This prototype was deployed and tested in public cloud Amazon EC2.

3.2. Some Experimental Results Explained

While testing, Farias *et al.* [10] covered most of the regression methods contained in *scikit-learn* library, to produce the predictive models along with *grid-search* hyperparameter optimization. The two methods that produced the most interesting results are: (i) *Linear Regression with Regularization* (LR) since it is a simple and an interpretable method; and (ii) *Gradient Boosting Machine* which is a *boosting* approach based on tree.

In statistics, the coefficient of determination, denoted R^2, is the proportion of the variance in the dependent variable that is predictable from the independent variable(s). This approach was applied in order to predict the mean response time (MRT) per second of all requests successfully executed by the system. Then, the results were compared when using linear feature (LF) or quadratic features (QF). LR obtained reasonably good R^2 score, whereas GBM obtained the highest R^2 scores in all cases. The results are shown in Table 1.

Table 1: Mean Response Time: R^2 score for linear and quadratic features [10]

	LR	GBM
LF	0.49	0.85
QF	0.74	0.83

The LR methods achieved an average R^2 score when using linear features (R^2 = 0.48). It suggests a non-linearity relationship of MRT metric and the system's parameters. When using quadratic features, LR methods improved the score to 0.74, which is an enhancement of 51% when compared to its linear equivalent.

In addition, Farias *et al.* [10] showed that GBM method captures highly non-linear relationships and performed well in both cases (linear and quadratic features). GBM performed better than LR with quadratic features because it captured other aspects of performance that LR with quadratic features could not consider. Their main conclusion is that GBM accurately predicts performance metrics considering concurrency and distribution aspects.

4. Conclusion

This review paper presents the concepts of NoSQL systems. Also, it has engaged in machine learning techniques for performance modeling of NoSQL databases. This approach is able to construct predictive models that predict performance metrics by capturing non-linear effects caused by the aspects of distribution and concurrency model. The literature review shows that a number of experiments have been carried out in order to evaluate this approach, where the target performance metric is the response time per second of the request successfully executed by the system.

References

1. Earley, S.: Data Virtualization and Digital Agility. IT Pro (September/October). IEEE Computer Society. (2016)
2. Bansel, A., Gonzalez–Velez, H., Chis, A.E.: Cloud-based NoSQL Data Migration. In 24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing. IEEE. (2016)
3. Gudivada, V.N., Rao, D., Raghavan, V.V.: NoSQL Systems for Big Data Management. IEEE 10th World Congress on Services. IEEE. (2014)
4. Bhogal, J., Choksi, I.: Handling Big Data using NoSQL. In 29th International Conference on Advanced Information Networking and Applications Workshops. IEEE. (2015)
5. Gessert, F., Ritter, N.: Scalable Data Management: NoSQL Data Stores in Research and Practice. ICDE Conference. IEEE. (2016)
6. Srivastava, P.P., Goyal, S., Kumar, A.: Analysis of Various NoSql Database. In 24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing. IEEE. (2016)
7. Lawrence, R.: Integration and Virtualization of Relational SQL and NoSQL Systems including MySQL and MongoDB. In International Conference on Computational Science and Computational Intelligence. IEEE. (2014)

8. Chen, L., Patel, S., Shen, H., Zhou, Z.: Profiling and Understanding Virtualization Overhead in Cloud. In 44th International Conference on Parallel Processing. IEEE (2015).
9. Zeng, J., Plale, B.: Workload-Aware Resource Reservation for Multi-Tenant NoSQL. International Conference on Cluster Computing. IEEE. (2015)
10. Farias, V.A.E. Flavio R.C.S., Maia, J.G.R., Gomes, J.P.P., Machado, J.C.: Machine Learning Approach for Cloud NoSQL Databases Performance Modeling. In 16th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing. IEEE. (2016)
11. Pedregosa, F. et al.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, Vol. 12, 2825–2830. (2011)
12. MongoDB. [Online]. Available: http://www.mongodb.com (current September 2017).

Project Planning in an Iterative-Incremental Software Development: Courseware Case Study

Ljubica Kazi[1], Dušan Mićić[1], Biljana Radulović[1], Dragica Radosav[1], Vesna Makitan[1], Eleonora Brtka[1], Dijana Karuović[1]
[1] University of Novi Sad, Technical faculty "Mihajlo Pupin" Zrenjanin, Djure Djakovica bb, 23000 Zrenjanin, Serbia
ljubica.kazi@gmail.com, dusan89vr@gmail.com, biljana.radulovic66@gmail.com, dolores023@open.telekom.rs, vesna@tfzr.uns.ac.rs, eleonorabrtka@gmail.com, dijanakaruovic@gmail.com

Abstract. Planning in software project management is one of the most important phases. It directs efforts and resources utilization, while estimating constraints and predicting issues. Since it is based on human judgements in estimation and predicting, planning is considered activity of a high risk. Therefore, it is necessary to use tools and methods that could lead to more precise and consistent results. Project planning is particularly specific in contemporary agile approach to software development, where iterative and incremental delivery is the key activity. This paper presents research results in project planning of iterative-incremental software development and a case study of iterative-incremental software development within the ISF@TFZR project, that is organized internally at University of Novi Sad, Technical Faculty "Mihajlo Pupin" Zrenjanin, Serbia. Particular example is presented with student's courseware software development project.
Keywords: project management, planning, agile, iterative, incremental, software development, courseware.

1. Introduction

Software "project management processes include processes for planning and estimating, resource management, measuring and controlling, leading, managing risk, managing stakeholders, and coordinating the primary, supporting, organizational, and cross-project processes of software development and maintenance projects." [1] Project's life cycle is usually separated into distinct phases or system components development activities. All project management process groups (initiating, planning, executing, monitoring and controlling, and closing)[2] could be repeated for each project's phase or subcomponent development activities group. Process initiation results in project charter, while planning main artifact is the project management plan, which [2], [3] should contain information about all ten aspects of project management knowledge areas. Software development process [3] is organized within software development life cycle (SDLC) and software product life cycle (SPLC) [1]. Well-known SDLCs [1] include the waterfall, incremental, and spiral models [4] plus various forms of agile software development [5]. "Agile models typically involve frequent demonstrations of working software to user representative, who directs development of the software in short iterative cycles that produce small increments of working, deliverable software". [1]

For software development within specific iterative-incremental, i.e. agile software development life cycle, in projects planning particularly important are activities of requirements prioritizing [6] and release planning [7]. "Delivering software in an incremental fashion implicitly reduces many of the risks associated with delivering large software projects

[7] "The complex nature of project management may require the use of repeated feedback loops for additional analysis." [2] Distinguishing SDLCs is partly based on requirements gathering and management methods. "Linear development models typically develop a complete set of software requirements, to the extent possible, during project initiation and planning. An incremental model produces successive increments of working, deliverable software, based on partitioning of the software requirements for increments implementation. Agile models define product scope and high-level features initially and facilitate evolution of software requirements during the project." [1]

This paper presents related work analysis and description of practical results in software project planning in iterative-incremental software development (IISD). The case study demonstrates the example of a student's courseware software development project that belongs to a broader ISF@TFZR project.

2. Related Research

One of the very important issues in software industry is acceptance and adopting of iterative-incremental approach to development, while moving from linear SDLC to agile practices. In [8], an interview-based case study is performed in Ericsson AB, Sweden in aim to investigate the perceived change when migrating from plan-driven to agile software development. Both approaches were identified with issues, but conclusion is that agile practices bring added values in absence of critical issues that are encountered in plan-driven development. Interview-based research is continued for large-scale software development within industrial practices and comparison of the case study with the state of the art [9]. Global software development is contemporary practice in software industry. Issues in using iterative and incremental processes in global software development are examined in [10]. The case study included five projects that are organized with distributed development teams and customer companies (average – three countries per project). The main practices in these projects were in: delivery synchronization, early design and code reviews, emphasis on communications, feature-based development, behavioral patterns, and frequent deliveries. Benefits gained in the case study projects: transparency of progress, instant feedback, and flexibility, ensuring joint understanding or requirements, avoiding "big bang" integration. One of the most important aspects in iterative-incremental software development planning is related to selection and prioritizing of requirements as well as to properly define iterations, by addressing the selected requirements to be considered for software functionality development within certain time-frame. Development period, for each iteration, could be with fixed length, usually a week or month [10], upon agreement in development team. Timeboxing[11] is approach that introduced fixed time box for iterations – each iteration has the same time period for all activities and what should be developed within an iteration is adjusted to fit the time box. Usually, time boxed iterations are executed in sequence, continuing work from previous iteration. In [11], the timeboxing is proposed to be used with pipelining concept, where timeboxed activities are executed concurrently, which leads to reduction in delivery time for product releases. Iterative-incremental software development affects overall quality of the

software product. Using defect analysis feedback could improve quality and productivity in iterative-incremental software development [12]. The research [12] shows the significance of constant monitoring and improvements in development process by having defect analysis feedback upon one iteration that could lead to less defects and better productivity in future iterations. Some of the defect causes from the example presented in [12] are related to: applying standards in semantic and coding standards (using, updating versions), redundant code, lack of understanding of use cases, existing code, object model and database, user interface and architecture issues. Within iterative-incremental software development planning, project economics (including effort estimation) is one of very important aspects. In [13], research is conducted in aim to determine does the total effort depend on the number of increments, what is optimal number of increments and how does the total development effort in incremental development relate to similar "integrated" development. The authors use COCOMO II model to answer these questions, but use assumptions and simplifications, such as: core software delivery + iterations, initial software design linearly evolve with new iterations, equal size increments, the same development team with all iterations etc. These simplifications lead to certain fixed parameters in equations that deal with software development effort calculations, but require further research with flexibility in previously mentioned assumptions.

3. Case Study

The case study in this paper presents results in project planning with iterative-incremental approach with example of courseware software development. The particular project of courseware software development is one of the students' projects within broader project which aims to empower students' professional knowledge and skills, by engagement within real-life projects related to improvement of university school (faculty) information system. The broader project is an internal project entitled "Professional empowerment of students within university school information system improvement" and it is conducted since year 2014 at University of Novi Sad, Technical faculty "Mihajlo Pupin" Zrenjanin (short name for the project: ISF@TFZR). Each student's project, from ISF@TFZR group of projects, is conducted in steps, as given at Table 1.

Table 1. Overall life cycle of students' projects within ISF@TFZR projects

Project Initiation and Software requirements specification	1st step – general project initiation: Students are informed about the general project and invited to attend the general project meeting, where the proposed topics (by the general project lead) are elaborated or students could propose their own software-related topics related to improvement of information system.
	2nd step – project scope and requirements elaboration: Students are given, for the proposed topics, the particular project's scope and initial requirements document and they attend personalized meetings with the particular project mentor, where requirements are explained in more detail

	and discussed with students and they could propose additional features.
Software design and Project planning	3rd step – project planning: Students prepare project proposal document that incorporates data from the project charter and project management plan (it does not include any models or detailed business process analysis or system design). The mentor refines the project proposal in consultations with the student. Non-functional requirements (including the software implementation technology) and development iterations are agreed and included in project proposal document.
	4th step – software design: If the software is recognized as business process-oriented, student creates business process model (BPM) and proposes software design, based on mapping with elements of BPM. According to methods from [14], functional software requirements i.e. software functions, categorized by user roles and development priorities (basis for iteration planning), are derived from primitive business processes, while conceptual data model is created based on business process data stores. Software design is documented with core UML diagrams: use case, component, deployment and data models (conceptual and physical data model and UML class diagram). The flexibility of the proposed design to the expected software enhancements (within the planned development iterations) is discussed and carefully adjusted in the software design.
Project Executing with Software construction	5th step - Iterative-incremental development of software: Each iteration ends with completely functional software ready for deployment (SQL script, source code, installation), tested with real-life data and documented with appropriate design models, user manual and technical documentation (installation and maintenance guide etc.). Iterations are defined with general time-boxing. ISF@TFZR project each year starts in March and ends in November -that is general project time-frame for each student's project. Most students take a general topic that could be implemented software within continuing projects during several years. Each project has internal iterations at week and month level. Primary focus in iteration planning is oriented to functionality modules delivery. An iteration could be part of a multiple-year project or each iteration could be organized as separate project. 1st iteration (project) – core functions that support software specification 2nd iteration (project) – software functions related to business process primitive process support ("need to have" software functions) 3rd iteration (project) – additional software functions that extend software functions related to business process primitive process ("nice to have" software functions) 4th iteration (project) – support for diversity of user roles, data validation, business rules application, interoperability with other applications within existing solution to business process support (export and import of data), business process automation.

Results of project initiation and planning for the student's project "Courseware web application development" are presented briefly with the most important elements.

Project initiating (Project charter) – The *scope of the project* is to develop a web application that could be integrated with existing web site of TFZR (www.tfzr.uns.ac.rs). The *business (social and market) need* for development of courseware web application arises from the need of professionals for constant knowledge and skills improvements, following the progress of contemporary technologies, methods and tools. It is common that university schools (such as MIT [15]) and specialized informal education institutions (such as [16]) organize additional courses related wide range of topics as additional education (web catalog of profession-related courses and providers is available at [17]). The *strategic plan* at TFZR is to provide formal and informal education and to keep education process aligned with knowledge requirements for business. TFZR is legally accredited for higher education and also providing informal education within additional courses. *Technical requirements* are related to the development of a web application as PHP or ASPX since these technologies are already supported at TFZR website hosting and new module should be integrated with existing solution).

Initial software requirements – The software that is to be implemented is recognized as business process-oriented. Initial functional requirements are defined upon textual description of well-known business process in courses organization: establishing a course, organizing teaching rooms and teaching staff, as well organizing course attendees' registration, attendance and examination. The initial user roles for the software could be: secretary (data input on teaching staff, teaching locations, courses and teaching/examination schedule), teaching staff member (data input of teaching topics, assignments and exam results), course attendee (browsing and filtering list of courses, data input on registration for course attending). All basic software functions could be extended with appropriate tabular view and filter, reports, update and delete functions etc.

Software design – Business process model (figure 1 – left part) is created within [18] according to previously mentioned business process description and mapped, by using the mapping table (figure 1 – right part), to software functions of first and second priority and precondition software functions. Data stores, which collect and store relevant data in the system, are decomposed to the set of entities in conceptual data model.

Project iterations scope for prototype iterative-incremental development – the example of courseware web application has been, so far, developed within two iterations (organized as two one-year ISF@TFZR projects) and the detailed project plan is ready for the third iteration.

- *First iteration:* The core functions were developed within BSc studies final exam work [18], where the set of initial requirements for software functions were derived from short description of business process and general analysis of available similar software. Figure 2. presents one of the resulting web pages from first iteration.
- *Second iteration* was conducted within MSc studies practical exam work within Subject "School information systems"[19] and it was based on detailed business process modeling and mapping of business process model elements to software design. Figure 3. presents the same web page from the Figure 2, enhanced with additional functionality and data.

- *Third iteration* was planned to be developed upon additional software requirements that arise from the analysis of existing solution of course-related data at TFZR website, as well as detailed analysis of other available solutions from this courseware software category. The complete set of requirements is organized within the project proposal document for IST@TFZR project for the year 2017, as continuation to previous efforts.

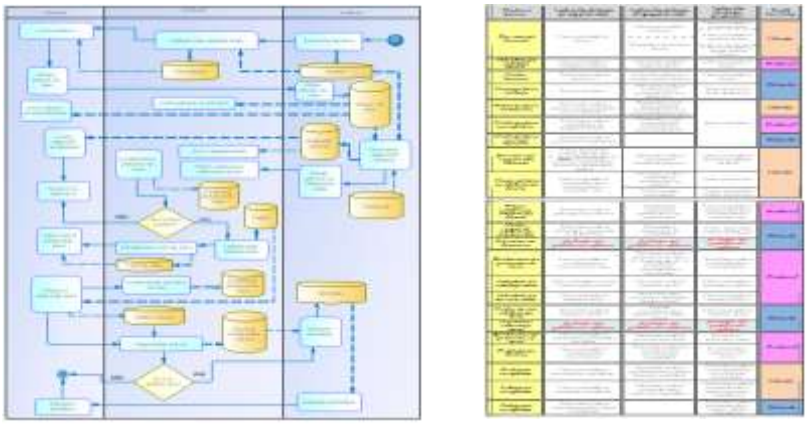


Figure 1. Business process model for organization of courses and table for mapping software primitive business processes to software functions [18]

Figure 2. List of courses page, First iteration version [19]

Figure 3. Improved list of courses page, Second iteration version [19]

Project management plan – Project charter is, for all iterations, equal in some parts, such as general scope and objectives („development of web application for course management"), business need, business process description, user-roles and non-functional requirements (integration issues related to technology, such as "existing web site hosting supports ASPX and PHP"). Every mayor development iteration, as separate one-year ISF@TFZR project, had the same general time-frame and weekly/monthly internal iterations. Regarding all other aspects of project management knowledge areas, for all iterations/projects values are the same, except for the project scope/requirements: stakeholders (university school management finds development of the courseware software necessary for organizing courses, which will enable raising financial resources of the institution), cost (the students work is not payed and all other costs for material are minimized to a single CD for final delivery of project results), quality metrics (software functions according to requirements and business process needs, software testing results, user interface design), human resources (single student is engaged in all project phases, from planning to delivery and testing), communications (student and menthor – via email and personal meetings), risks (factors: technology knowledge, unrealistic estimation of scope regarding available development time, because of other students work engagements in study process), procurement (there is no external dependency on other companies that should provide part of solution or any material). Activities for each iteration/project were performed with the same sequence, including: project initiation, project planning, software design, construction, testing and documenting (user manual). The scope for each iteration/project is previously described.

4. Conclusion

There are many important issues that are related to iterative-incremental software development methodology, which encourage research and new methods development. These issues include adopting iterative –incremental approach to development companies that previously practiced linear development and in global software development with synchronization and communication difficulties, scheduling with timeboxing, using defect analysis feedback for process improvement and project economics with effort and cost estimations.

This paper focused on the problem of project planning in software development projects based on iterative-incremental paradigm. The case study, related to student's software of developing courseware web site, is described. The case study demonstrates iteration planning results and other important aspects of knowledge areas within general project management, that could be applied for the software development. In this case study, the same software ("courseware web site") was developed during three one-year lasting projects. Each project continued from the previous one, by adding new software functionality or enhancing existing ones. Therefore it is presented that a software could be developed as single multiple-year project or as a portfolio of multiple projects related to the same topic. With this case study, it has been shown that the approach of multiple one-year projects lead to successful solutions. It

is particularly suitable for students' projects, where their activities are combined with other learning engagements and study assignments.

References

1. Guide to the Software Engineering Body of Knowledge SWEBOK v3.0, IEEE Computer Society, (2014).
2. A Guide to the Project Management Body of Knowledge PMBOK, Fifth Edition, Project Management Institute, (2013)
3. Pfleeger S.L, Atlee J.M: Software Engineering Theory and Practice, Pearson Education, 2006.
4. Boehm B: A Spiral Model of Software Development and Enhancement, Journal IEEE Computer, 61-72. (1988)
5. AGILE MANIFESTO, http://agilemanifesto.org/ [Accessed 13.9.2017.]
6. Firesmith D: Prioritizing requirements. Journal of object technology. Vol. 3, No.8, (2004)
7. Ruhe G, Greer D: Quantitative Studies in Software Release Planning under Risk and Resource Constraints. Proceedings of the International Symposium on Empirical Software Engineering (ISESE '03), 1-10, (2003)
8. Petersen K, Wohlin C: The effect of moving from a plan-driven to an incremental software development approach with agile practices, an industrial case study. Journal Empirical Software Engineering, Springer, Vol 15, 654-693, (2010)
9. Petersen K, Wohlin C: A comparison of issues and advantages in agile and incremental development between state of the art and an industrial case, Journal of Systems and Software, 1-15, (2009)
10. Paasivaara M, Lassenius C: Using Iterative and Incremental Processes in Global Software Development, Proceedings of the 3rd International Workshop on Global Software Development, Edinburgh, Scotland, 42-47, (2004)
11. Jalote P, Palit A, Kurien P: The Timeboxing process model for iterative software development, Journal of Systems and Software, Vol. 70, No 1–2, 117-127, (2004)
12. Jalote P, Agrawal N: Using defect analysis feedback for improving quality and productivity in iterative software development. 3rd International Conference on Information and Communications Technology, Enabling Technologies for the New Knowledge Society, 1-11,(2005)
13. Benediktsson O, Dalcher D, Reed K, Woodman M: COCOMO-Based effort estimation for iterative and incremental software development, Software Quality Journal, Vol. 11, 265-281, (2003)
14. Kazi Lj, Radulovic B, Radosav D, Bhatt M, Grmusa N, Stiklica N: "Business Process Model and Elements Of Software Design: The Mapping Approach", "International conference on Applied Internet and Information Technologies", ICAIIT2012, Zrenjanin, 17-20 (2012)
15. MIT Open courseware, https://ocw.mit.edu/index.htm [Accessed 4.9.2017.]
16. Academy Oxford courses, http://www.akademijaoxford.com/strucni_kursevi_i_obuke_srbija.php[Accessed 4.9.2017.]
17. Web catalog of courses and providers, ww.kursevi.com [Accessed 4.9.2017.]
18. Micic D: Developing web application for faculty courses organization, BSc final work, University of Novi Sad, Technical faculty "Mihajlo Pupin" Zrenjanin, (2016)
19. Micic D: Improvement of web application for faculty courses organization, Practical Exam Work at master studies, University of Novi Sad, Technical faculty "Mihajlo Pupin" Zrenjanin, (2017)

Investigating Organizational Issues in a Micro Software Company: A Thematic Data Analysis

Karin-Mira Sisak[1], Dalibor Dobrilovic[2], Zeljko Stojanov[2]

[1]West Regional Development Agency, Timisoara, Romania
mirasisak26@gmail.com
[2]University of Novi Sad, technical faculty "Mihajlo Pupin", Zrenjanin, Serbia
dalibor.dobrilovic@uns.ac.rs, zeljko.stojanov@uns.ac.rs

Abstract. Software companies are considered as complex socio-technical organizations, which requires considering technical, organizational and social issues. However, literature mostly deals with technical issues, although organizational and social issues strongly influence practice in software companies. The problem becomes even more important in small software companies that have several constraints related to resources, funding and staffing. This paper presents a segment from a case study organized in a micro software company aimed at investigating maintenance practice. This segment relates to organizational issues that influence software maintenance technical processes. Thematic framework that reflects organizational issues is developed by using inductive thematic analysis. The paper concludes with the discussion of research implications, limitations and further research directions.

Keywords: organizational issues, business policy, software company, thematic data analysis.

1. Introduction

Organizational and social issues have been recognized as important factors for business performance of software organizations, which resulted in increased number of studies focusing on these issues in software engineering practice. These trends resulted with inclusion of chapters *Software Engineering Professional Practice* and *Software Engineering Economics* in *Guide to the Software Engineering Body of Knowledge* book published in 2014 [1]. In one of the most influential book considering organizational issues in software engineering *The Mythical Man Month* [2], Fred Brook suggested that bad communication between people and teams is one of the main reasons for software projects' failures, which can be overcome in well-structured software organizations. Moreover, improvements of software processes and practice should encompass human, organizational and technological factors [3].

Inquiry of organizational and social issues needs different approach compared to technical and technological issues that dominate in software engineering practice. Successful software engineering practice require considering both aspects of the practice. Even more, some authors argued that organizational issues are more important than technical ones [4]. In a comprehensive literature review of various situational

circumstances affecting software engineering practice, Clarke and O'Connor [5] classified factors in several categories that reflect different issues in software organizations. The main categories related to organizational issues are: (1) Personnel - characteristics of people involved in software engineering practice, (2) Organization - profile of organization where the practice is situated, (3) Operation - Consideration and constraints enabling operations in an organization, (4) Management - Characteristics of management team and structure in software organizations, and (5) Business - strategic and tactic business considerations.

The main challenge for organizations in competitive market is how to transform the implicit, fragmentary and private knowledge of individuals or groups into valuable intellectual assets that will increase organizational overall business performance and competitiveness [6]. For small organizations, knowledge, abilities and skills of staff are the most important and renewable resources that should be preserved for future use. Majority of knowledge in small enterprises is kept in the heads of the key employees rather than in physical storage, which presents the permanent danger for losing knowledge if they leave the company [7]. Therefore, knowledge identification is an initial and the most important step in adopting knowledge management (KM) in micro and small organizations, deserving more attention by research community related to developing methods for capturing organizational knowledge [8]. The main constraint of small organizations in adopting and implementing KM in their practice relates to the lack of resources and capacities (human, financial, labor, equipment), making their approaches to KM informal [9]. Software industry consists of mostly small organizations that share similar problems in the fields of process improvement and KM with other small organizations, such as constraints in funding, people and resources. In addition, they rarely adopt best practice proposals and standards, but rather develop their own methods for managing development and maintenance activities [10]. The main problem in software maintenance is the lack of various types of knowledge, as well as methods for capturing and sharing knowledge [11]. Based on the findings from empirical study, Jantunen [12] concluded that existing software engineering approaches need to be expanded with approaches that are more social by nature and are more tolerant of improvisation.

After investigating software maintenance practice in software companies located in Banat region in Romania and Serbia, it becomes evident that software maintenance is the area that needs further investigation and improvements, especially in the areas such as process improvement and knowledge management [13]. In line with these observations, a project for investigating and improving practice in a local software company was organized. As a part of this project, identification and systematization of knowledge used in software maintenance practice was conducted. This paper presents a segment of this research oriented towards organizational issues in the company. The aim of this paper is to present knowledge related to organizational issues rendered as a thematic framework that can be used as basis for structuring knowledge in the company.

2. Related Work

This section outlines studies dealing with different organizational issues affecting software engineering practice. These studies are selected to cover different phases in software life cycle (e.g. requirements, development, or maintenance) and different organizational structures in software organizations.

Nagappan et al. [14] investigated relationship between organizational structure and software quality. The authors proposed a set of eight measures for quantifying organizational complexity: Number of engineers, Number of ex-engineers, Edit frequency, Depth of master ownership, Percentage of organization contributing to development, Level of organizational code ownership, Overall organization ownership, and Organization intersection factor. Based on proposed measures the authors investigated the effect of organization structure on software quality (predicting failure-proneness in Windows Vista).

Belfo [15] conducted literature research on management practice in requirements engineering (RE) based on Orlikowski multi-dimensional view that include peoples, organization and technology [16]. The author discussed managerial questions in RE, such as clearness, completeness, correctness, understandability, verifiability and validity by providing relevant questions from organizational perspective. For each question the author suggested literature sources.

Tamburri et al. [17] investigated the state of the art in Organizational Social Structures (OSS) in the context of software engineering. OSS can be seen as a set of interactions, relations and social arrangements between individuals working together within some organizational context. The authors conducted a systematic literature review (SLR) of OSSs using grounded theory, and based on the findings proposed guidelines for further research related to socio organizational issues in software organizations.

Ulziit et al. [18] presented a conceptual framework that explains challenges and solutions for management in global software maintenance. The focus of research is on communication, coordination and control, which appear as the most significant in global distributed working environment. The main challenges identified by applying grounded theory method on data gathered through literature review and interviews are: quality management, scope management, change management, communication management and risk management. For these types of challenges the authors discussed possible solutions.

3. Case Study

Case study was organized in a micro software company oriented towards developing and maintaining software products for local clients in Serbia. The study was implemented as a part of long term project related to process improvement in the company. The research was conducted as a joint work of researchers and company employees, with support of experts from regional development organizations and business incubators.

The company has seven employees, six programmers and one technical secretary. Business software solutions were developed for local market, while the majority of activities relate to maintenance that consumes the majority of time and costs in the company [19]. Identification and specification of knowledge was implemented by using basic principles of organizational learning [20].

3.1. Research Methods

Knowledge identification and specification was implemented through development of a context specific method called *Lightweight Inductive Method for Knowledge Identification and Systematization* (*LIM4KIS*) [21]. The method assumes joint work of researchers and organizational staff for identifying the real knowledge in an organization, while it has foundations in well established scientific research methods.

Data were collected by using different social science methods, which were recently adopted in empirical studies in software engineering. These methods include semi-structured in-depth interviews with open-ended questions, practice observation, field note taking and organizing working meetings in the form of groups with focused discussions [22]. In addition, tasks related data were extracted from the company internal repository of tasks by using purposely prepared SQL scripts [19]. Development of a thematic framework was based on inductive thematic analysis proposed by Braun and Clarke [23].

3.2. Research Process

Research process includes three main phases: planning, implementation and reporting. Planning activities include joint work of the company management and researchers. Implementation phase assumes active participation of all company employees, which ensures identification of the real practice in the company. Reporting was duty of researchers, while company management validated all reports.

The main phase is implementation of proposed LIM4KIS method in the company context. However, used thematic analysis method was slightly modified during the implementation, since qualitative research methods that dominate in this research allow adjustments of research design during the research process [24]. The overall presentation of LIM4KISS method is presented in Fig. 1, where data sources, data analysis methods, some activities specific for LIM4KIS, such as feedback sessions, and outputs can be clearly distinguished.

The central parts in the LIM4KIS method are feedback sessions, which are organized as working meetings in the company. The researchers and the company staff participated in feedback sessions based on the current state of the research process. Feedback sessions were organized several times during the research, since it has been recognized that feedback is inevitable part of any improvement projects in software engineering [25]. The discussions during the sessions were tape recorded, transcribed and used in thematic data analysis for developing thematic framework with knowledge. In addition, during the sessions, situations that require additional data to be collected and analyzed to justify current findings were identified.

Due to the complexity and dynamics of software maintenance practice in the company, thematic data analysis approach proposed by Braun and Clarke [23] was slightly modified, reflecting flexibility in a study design [24]. The modification includes identification of thematic areas related to specific aspects of software maintenance practice in the company, and after that classification of initial codes and developed themes into identified thematic areas. Thematic data analysis included the following phases: (1) Familiarizing with the data, (2) Generating initial codes and identification of thematic areas, (3) Searching for themes and clarification of thematic areas, (4) Reviewing themes and thematic areas, (5) Defining and naming themes and thematic areas, and (6) Preparing a report with developed thematic framework. This paper presents only a segment of the thematic framework (thematic area) related to organizational issues in software maintenance practice in the company.

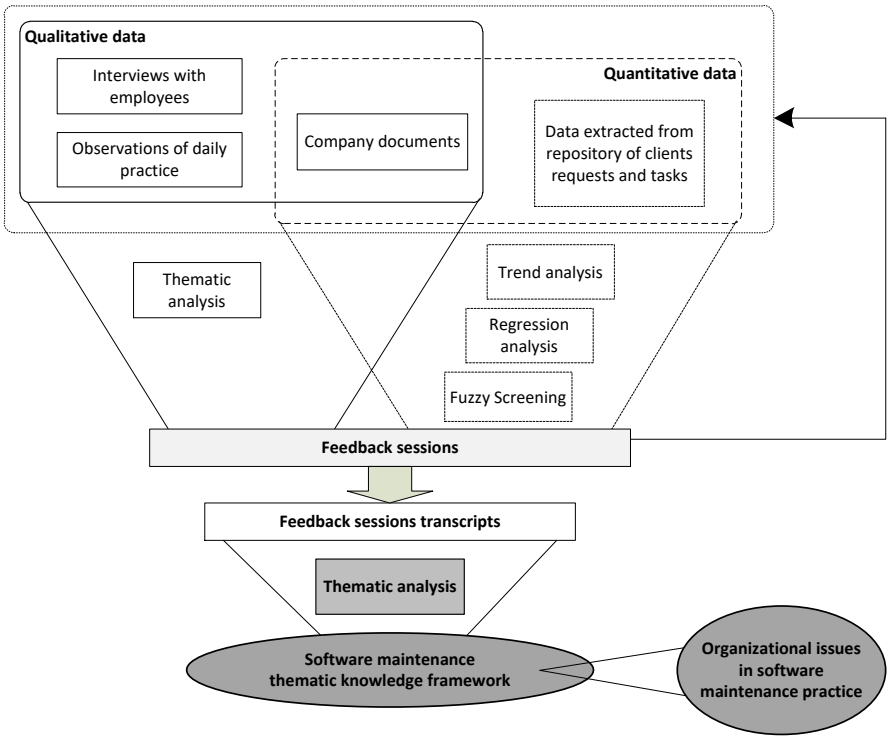


Fig. 1. Overall presentation of LIM4KIS method

Since qualitative data analysis depends on researcher who carries out the codification [26], the objectivity of data analysis in the study was increased by involving the company staff in data analysis by organizing feedback sessions in the company.

3.3. Findings

The research findings were developed as a thematic framework. Framework is grounded in the evidence about the practice in the company, including context, individuals and actions. According to the typology of qualitative findings proposed by Sandelowski and Barroso [27], these findings can be classified as conceptual/thematic description, since they are presented as a network with identified themes and sub-themes.

Through the years of experience at the local market, the company has developed its specific business policy and established an internal organization. These are mostly influenced by the preferences and knowledge of three leading programmers, and modified through the extensive business activities and the company growth. Fig. 2 presents themes related to organizational issues, mostly related to software maintenance practice.

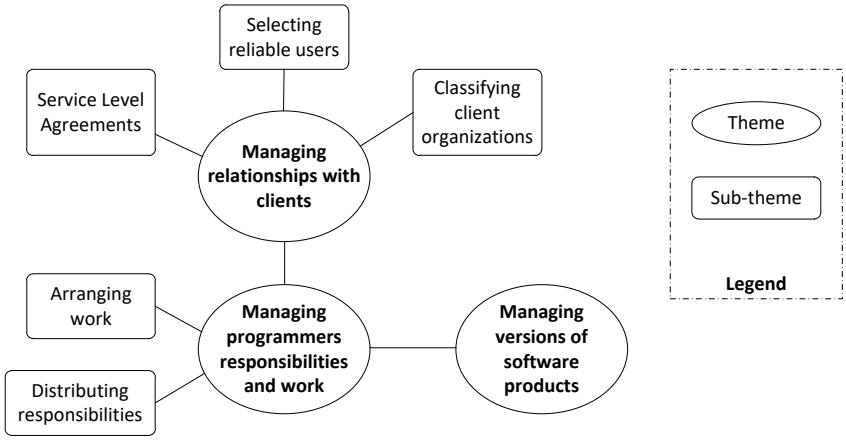


Fig. 2. Identified themes related to organizational issues

The organizational issues relate to establishing and maintaining reliable relationships with client organizations, as well as managing the internal resources to ensure efficient maintenance processes. The most important internal resources in software maintenance are programmers and software products. The company organization relates to everyday work that should ensure completing all activities with the optimal use of the resources and the maximum quality of the services and products. Identified themes are: (1) Managing relationships with clients, (2) Managing programmers' responsibilities and work, and (3) Managing versions of software products.

Managing Relationships with Clients

The clients are very important for the company since they are the main source of profit. This strongly emphasizes the importance of managing the relationships with them. These relationships can be distinguished as relationships between: (1) the company and client organizations, which are defined in Service Level Agreement (SLA) documents,

and (2) individual programmers and software users from the clients' organizations. Identified sub-themes are:

- *Service Level Agreement.* The highest level of managing relationships with the clients is at organizational level, which is implemented through SLAs. These documents arrange the level of maintenance services that the company provides to each specific client organization.
- *Classifying client organizations*. The classification of client organizations enables more efficient handling of maintenance requests (MRs). In addition to the classification based on SLAs, the company classifies clients based on the client field in economy and based on financial issues related to paying maintenance services.
- *Selecting reliable users*. Programmers gained experience with different users from clients' organizations. Due to several problems with users in communicating MRs, the company identified reliable users in each client organization who are allowed to submit MRs.

Managing Programmers' Responsibilities and Work

Organizing programmers' work on solving MRs is essential for overall business performance of the company, since maintenance tasks consume majority of working activities in the company. Identified sub-themes are:

- *Distributing responsibilities*. This sub-theme relates to organizing and planning work in the company in a way that each software application has more than one assigned programmer that can solve MRs. Properly distributed responsibilities ensure faster and more reliable solving of MRs.
- *Arranging work*. This sub-theme relates to solving particular situations related to MRs, which requires programmers' organizational skills. Arranging work may include self organization based on previously assigned MRs and other tasks, asking for help from other programmers, or forwarding MRs to other more competent programmers.

Managing Versions of Software Products

The company maintains over 30 software products used by over 100 client organizations from different business fields. Different clients' organizations use different versions of the same software product, and it becomes important to effectively manage distribution of different versions of software to different client organizations. In addition to MRs submitted by users, development of new versions of software products is driven by the country regulative.

3.4. Limitations of the Study

The validity and rigor of qualitative research studies are based on ensuring that trustworthiness criteria are respected [28]. The credibility, or internal validity, of the

study was ensured through careful application of inductive thematic analysis, rich descriptions of the context, process and findings, and triangulation of data sources.

The authors are aware that transferability of the research findings is also limitation of this study and threat to trustworthiness. However, the authors do not claim transferability of the research findings, but believe that other researchers can find guidelines how to develop methods similar to LIM4KIS and implement them in variety of organizational settings for inquiring specific segments of the practice.

4. Benefits for the Company

Since the company management and all employees participated in the research, the company expected clear benefits. The first benefit relate to creation of a base of knowledge related to organizational issues in the company, which can be used for planning activities and as base for creating database of company specific knowledge for young programmers.

The second benefit relates to introduction of knowledge management and organizational learning practice (and culture) in every day practice, which can be treated as basis for further process improvement projects.

5. Conclusions

Organizational issues are essential for overall business performance in software organizations. Identification of these issues in the real organizational context helps in identifying problems and improving practice. This paper presents an approach for identifying organizational issues in software maintenance practice based on inductive thematic analysis, as well as findings organized in a thematic framework that reflect the real practice.

Further work will be directed in several directions. The first one is development of a software tool for managing knowledge organized in thematic frameworks. This software toll will enable sharing of existing knowledge and adding new knowledge discovered in the practice. Implementation of this approach for knowledge identification and systematization in other software companies is another promising research direction.

Acknowledgement

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, 2011-2017.

References

1. Bourque, P., Fairley, R. E. (Editors): Guide to the Software Engineering Body of Knowledge, Version 3.0, SWEBOK. IEEE. (2014).
2. Brooks, F. P.: The Mythical Man-Month, Anniversary Edition: Addison-Wesley Publishing Company. (1995).
3. Perry, D. E., Staudenmayer, N., Votta, L. G.: People, Organizations, and Process Improvement. IEEE Software, Vol. 11, No. 4,36-45. (1994).
4. Dybå, T.: Enabling Software Process Improvement: An Investigation of the Importance of Organizational Issues. Empirical Software Engineering, Vol. 7, N. 4, 387-390. (2002).
5. Clarke, P., O'Connor, R. V.: The situational factors that affect the software development process: Towards a comprehensive reference framework. Information and Software Technology, Vol. 54, no. 5, 433-447. (2012).
6. Ho, C.-F., Hsieh, P.-H., Hung, W.-H.: Enablers and processes for effective knowledge management. Industrial Management & Data Systems, Vol. 114, No. 5, 734-754. (2014).
7. Alvarez, I., Cilleruelo, E., Zamanillo, I.: Is formality in knowledge management practices related to the size of organizations? the Basque case. Human Factors and Ergonomics in Manufacturing & Service Industries, Vol. 26, No. 1, 127-144. (2016).
8. Hall, S. J., De Raffaele, C.: Corporate Amnesia in the Micro Business Environment. In: Proceedings of the 2013 Federated Conference on Computer Science and Information Systems (pp. 1235–1239), Kraków, Poland. (2013).
9. Liao, Y., Barnes, J.: Knowledge acquisition and product innovation flexibility in SMEs. Business Process Management Journal, Vol. 21, No. 6, 1257-1278. (2015).
10. Pino, F.J., Ruiz, F., Garc´a, F., Piattini, M.: A software maintenance methodology for small organizations: Agile MANTEMA. Journal of Software: Evolution and Process, Vol. 24, No. 8, 851–876 (2012).
11. Anquetil, N., de Oliveira, K.M., de Sousa, K.D., Dias, M.G.B.: Software maintenance seen as a knowledge management issue. Information and Software Technology, Vol. 49, No. 5, 515–529. (2007).
12. Jantunen, S.: Exploring software engineering practices in small and medium-sized organizations. In: Proceedings of the 2010 ICSE Workshop on Cooperative and Human Aspects of Software Engineering (CHASE '10), pp. 96-101. (2010).
13. Stojanov, Z.: Using Qualitative Research to Explore Automation Level of Software Change Request Process: A Study on Very Small Software Companies. Scientific Bulletin of The "Politehnica" University of Timişoara, Transactions on Automatic Control and Computer Science, Vol. 57(71), No. 1, 31-40. (2012).
14. Nagappan, N., Murphy, B., Basili, V.: The influence of organizational structure on software quality. In: proceedings of the ACM/IEEE 30th International Conference on Software Engineering, pp. 521-530. Leipzig, Germany. (2008).
15. Belfo, F.: People, Organizational and Technological Dimensions of Software Requirements Specification. Procedia Technology, Vol. 5, 310-318. (2012).
16. Orlikowski, W. J.: The Duality of Technology: Rethinking the Concept of Technology in Organizations. Organization Science, Vol. 3, No. 3, 398-427. (1992)
17. Tamburri, D. A., Lago, P., van Vliet, H.: Organizational social structures for software engineering. ACM Computing Surveys, Vol. 46, No. 1, Article No. 3. (2013).
18. Ulziit, B., Warraich, Z. A., Gencel, C., Peterse, K.: A conceptual framework of challenges and solutions for managing global software maintenance. Journal of Software: Evolution and Process, Vol. 27, No. 10, 763-792. (2015).
19. Stojanov, Z., Dobrilovic, D., Stojanov, J.: 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, 59-74. (2013).

20. Argote, L.: Organizational Learning: Creating, Retaining and Transferring Knowledge, Second Edition. Springer US. New York, USA. (2013).
21. Stojanov, Z., Dobrilovic, D., Zarkov, A.: Lightweight inductive method for Knowledge Identification and Systematization - LIM4KIS. Technical report No 004-2015. Technical faculty "Mihajlo Pupin"Zrenjanin, University of Novi Sad, Serbia. [in Serbian] (2015).
22. L. M. (Editor): The SAGE Encyclopedia of Qualitative Research Methods. SAGE Publications. Thousand Oaks, CA, USA. (2008).
23. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qualitative Research in Psychology, Vol. 3, No. 2, 77-101. (2006).
24. Marshall, C., Rossman, G. B.: Designing Qualitative Research, Fifth Edition. SAGE Publications. Thousand Oaks, CA, USA. (2011).
25. Stojanov, Z., Dobrilovic, D.: The Role of Feedback in Software Process Assessment. In: Mehdi Khosrow-Pour (editor), Encyclopedia of Information Science and Technology, Fourth Edition, pp. 7514-7524, chapter 654. IGI Global. Hershey, PA, USA. (2017).
26. Moral, C., De Antonio, A., Ferre, X., Lara, G.: A coding system for qualitative studies of the information-seeking process in computer science research. Information Research, Vol. 20, No. 4, paper 699. (2015).
27. Sandelowski, M., Barroso, J.: Classifying the findings in qualitative studies. Qualitative Health Research, Vol. 13, No. 7, 905–923. (2003).
28. Schwandt, T. A., Lincoln, Y. S., Guba, E. G.: Judging interpretations: But is it rigorous? trustworthiness and authenticity in naturalistic evaluation. New Directions for Evaluation, Vol. 114, 11-25. (2007).

Legal Protection of the Computer Program:Copyright or Patent Protection

Nadezda Ljubojev[1], Dragica Ivin[1]

[1] University of Novi Sad, Technical Faculty "Mihajlo Pupin", Đure Đakovića bb, 23000 Zrenjanin, Serbia
nadezdaljubojev@gmail.com, ivin_bd@hotmail.com.

Abstract. The development of e-trade and technological society based on the Internet affects the expansion of production and the usage of computer programs so the issue of appropriate legal protection of computer programs is still present. In this paper, the authors analyze the system of the computer program protection both in the laws of European Union and the Republic of Serbia. The aim of the paper is to give an analysis of copyright and patent protection of the computer program, especially in relation to the process of harmonization of Serbian law during the EU integration process. Taking into account that both copyright and patent protection of the computer program exist in practice in the same time, the emphasis is on their relationship as well as on the possibility of future development.

Keywords: protection of the computer program, copyright, patent right.

1. Introduction

Digital technology has become an integral part of everyday life on the global level. This very fact suggests the great significance of the computer program and how important the issue of determination of adequate forms of its legal protection is.

Software industry needs legal protection for its further development. Relevant regulations should protect huge investments in hardware and software production and encourage even more investments as well. However, besides giving legal protection to the authors of computer programs from illegal usage of their programs, it is often necessary to regulate the relations between the authors and the users by law. There is also the need for a regulation of computer program transfer.

Regarding the importance of the electronic trade and development of technological society based on the Internet, the real expansion of production and use of computer programs is still expected, so the issue of their adequate legal protection on the European continent is very up-to-date. Because of the economic influence of the software industry in the United States (US) and their expressed efforts to attribute patent protection to the computer program as well, there is a doubt in justification of the copyright protection of the computer program in Europe, too.

Disagreement on legal protection can have negative effects on investment decisions and free flow of goods on the market. Competitive environment for innovations differs considerably depending on the fact whether a computer program is protected by the copyright or the patent. [1]

2. Copyright or Patent Protection of the Computer Program?

The issue of legal protection of computer programs in international and comparative law is not disputed, because of the fact that they represent an intellectual creation. There is a commitment that legal protection of computer programs should be within a specific branch of law – *intellectual property.* The question which is disputable is in what field of law on intellectual property legal protection should be provided – whether it is on the basis of patent rights regulations or on the ground of provisions of the copyright law, in order to establish a legal framework for exploitation of computer programs.

Copyright protection of computer programs can be suitable for software developers, because a detection of the program in the source code is not necessary for obtaining legal protection. That way, unlike the program in object code which is placed on the market, the source code remains a business secret of its creator.[2] Besides, copyright protection lasts for the life of the author and fifty (Berne Convention for the Protection of Literary and Artistic Works, 1971)[3] or seventy years after his death (Agreement on Trade-Related Aspects of Intellectual Property -TRIPS).[4]

Patent rights are not entitled automatically, by mere deed of inventing or public announcement. These rights are entitled only after the conducted administrative procedure for patent recognition obtained by filling the application form. The rights are valid from the filling date of the application to the institution relevant for the recognition of the patent rights. The patent is issued only when an invention satisfies the defined conditions of patentability. In order to be protected by the patent, an invention must represent a technical solution of a certain problem which is new, applicable and which is a result of creative work.

Namely, patent provides the protection for an invention within a patent claim. Therefore, the titular of a patent for computer applicable invention has the right to forbid to all third parties the use of patented invention in the volume determined by the patent application.

On the other hand, a complex nature of the computer program opens the question of protection by patent. The source and object code, as forms of expression, are the objects of copyright protection. However, activation of the computer program produces a technical result that often has an industrial application. Considering this fact, copyright concept of protection is being questioned not only by the American, but also by the European legal theory and practice. [5]

The attitude that computer programs represent author's work is protected by copyright provisions, overpowered after a comprehensive discussion.[6] This standpoint is expressed in the Provision of the Article 10 (1) of Agreement on Trade-Related Aspects of Intellectual Property Rights-TRIPS, according to which "*computer programs, whether in source or object code, shall be protected as literary works under the Berne Convention for the Protection of Literary and Artistic Works (1971)*". (Berne Convention for the Protection of Literary and Artistic Works amended in Paris on 4th May 1896, changed in Berlin on 13th November 1903, supplemented in Bern on 20th March 1914, changed in Rome on 2th June 1928, Brussels on 26 th 1948, Stockholm on 14 th July, 1967, and in Paris, 24 th July 1971).

However, that does not mean that legal protection of computer programs is uniformed in the comparative law.

3. The Copyright Protection of Computer Program

The Republic of Serbia (RS) harmonizes its laws in the field of intellectual property with EU *aquis communautaire*. By signing Stabilization and Association Agreement with EU, the RS has taken over the obligation to guarantee the volume of intellectual property protection similar to the volume existing in EU, including the instruments for the implementation of these laws.

The copyright protection of computer program in the Serbian law is in accordance with the regulations of the Directive on the legal protection of computer programs (Directive 91/250/EEC) in 1991, which represents a primary act regulation of the issue of the computer program's protection in the EU and it is used by the member states to protect computer programs with copyright, same as literary works, in terms of the Bern Convention (1971).[7] (Art. 1 (1) Directive 91/250/EEC)

The Directive 91/250/EEC was issued by the Council of the European Committees with the obligation of its applying in all of the member states of the European Economic Community starting from 1st January, 1993. It was composed in Brussels in May, 1991. By adopting the Directive on the legal protection of computer programs Directive 91/250/EEC in 1991, it was confirmed that computer programs were considered as author's literary works. This view was later also confirmed in the TRIPS[8], as well as in the Patent Law Treaty [9], which recognized the protection of computer programs as they are literary works.

The computer program is defined as "a series of instructions that are, upon their fixing on the machine-readable material carrier, capable of affecting the machine for information processing so it can express, perform or achieve a specified function, task or result".[10] The term software needs to be differed from the similar concept of the computer program, because the latter is a narrower term.[11] In theory, the term software is used often and it includes "the computer program and procedures, possible supporting documentation and data referring to the functioning of the computer system". [12] According to its purpose, software can be divided most frequently in two groups: system and application software, although from the aspect of legal protection this classification is not significant. The differences between them can affect certain issues in legal transactions. Speaking from the market aspect, the application software is most significant because it is intended for end-users in solving some business, scientific or industrial issues"[13] whose functioning is dependent on the operation system or service programs. [14]

The computer program expressed in any form is protected by copyright. In that way, copyright protects the titular from direct copying of the source or object code, but that protection does not apply on those cases representing different ways of expressing the same idea or principle. However, the titular may be called for a patent infringement when a computer application invention ,whose source or object code is different, has the same effect although, according to copyright law, this independently created program should also be the object of protection.

The computer program form includes: source code (when it is expressed in some of the program languages), object code (in form of binary numbers, i.e. in the machine language) and executable code (electronic digital inscription on the body carrier: magnetic tape, chip, CD). [11] "*Ideas and principles which underlie any element of a*

computer program, including those which underlie its interfaces, are not protected by copyright under this Directive". (Art. 1 (2) Directive 91/250/EEC) The computer program will enjoy copyright protection if it fulfills the conditions prescribed for the protection of the author's work. In order to be characterized as an authors work protected by law, a work needs to be the result of original human creativity expressed in form appropriate for that kind of work. No other criteria is applied for determining if a work falls under this protection (Art. 1 (2) and (3) Directive 91/250/EEC), because it is considered that the criteria used for evaluating the originality of a computer program should not contain any estimation of quality or aesthetic value of the program.

In accordance with the Directive 91/250/EEC, author of the computer program can be an individual or a group of individuals (co-authors). The right of protection belongs to the author or to co-authors, if more individuals participated in making of the computer program. If the computer program is created while the author is employed, employer is the holder of all exclusive authorizations on using the program, unless it is contracted differently between the employer and employee. .

The right holder has exclusive authorizations to allow or forbid permanent or temporary use of the program, in general or partly, in every way and by use of any technical means. It is defined, by the Directive 91/250/EEC, that use of the program means copying, storing in a computer or some other body carrier, preparing the program for work, adapting, adjusting or doing any other program change, as well as every other form of placing computer program on the market, including renting of the program. By national legislation, the member states are obliged to provide efficient means, in case of violation of rights of the author of the computer program.

Significant limits of the computer program's author are provided by the Directive 91/250/EEC, in favor of the program's lawful owner.

3.1 The Copyright Protection of the Computer Program in Serbian Law

In Serbian regulations on copyright, computer program as an author's work is regulated for the first time by the Law on the Amendments of the Law on Copyright from 1990. The Law on Copyright and Related Rights from 1998, ("Official Gazette", No. 24/1998), which followed, accepted the same concept by explicitly providing that computer program is on the list of author's works. (Art. 2 (2) (10)) The same law from 2004 ("Official Gazette", No. 61/2004), also did not bring essential changes, as it, according to international regulations, subsumes computer programs under written works, (Art 2 (2.) (1)) which is also provided by the Law on Copyright and Related Rights from 2009 ("Official Gazette", No. 104/2009 and No. 99/2011) (Art. 2 (2) (1)) (hereinafter: LCRR) LCRR anticipates computer programs as one of the author's works that are listed exemplary in the category of written works.

Considering the specifics of the computer program as an intellectual good, consistent application of provisions on author's literary works is not justified, and the rights of the author of the computer program are limited, in comparison with authors of other works.

Speaking about the computer program, it should be emphasized that suspension of copyright rights is especially regulated. Namely, the legislator has predicted that a person who legally obtained a copy of the computer program may, for his/her own purposeful usage of the program and without a permission of the author and also without paying the author's fee: store the program in computer's memory and put it into

operation; correct program errors and make other necessary changes in the program that are in accordance with its purpose unless otherwise provided by the contract; make a backup copy of the program on a permenent computer carrier; make a decompilation of the program exclusively due to obtaining necessary data for achieving interoperability of the program with another independently made program or certain computer equipment, under the condition that the data in question were unavailable and that the decompilation was limited only to the part of the program necessary for achieving interoperability (Art. 47 parag. 1 LCRR). Storing the program in computer's memory and putting it in operation represent in fact the activities of multiplication of the author's work implied by the exclusive authority of the computer's program author (Art. 47 parag. 1 LCRR). However, these activities are necessary so as a person who legally obtained a copy of the computer program could use it. The legislator decided to limit the afore mentioned authorities for this reason. Technical characteristics of the computer program imply that its making and usage require the use of a computer (machine), so software errors are frequent. Every unauthorized intervention would be considered as a change of the original work which means that it would violate an author's moral right on the protection of his/her work integrity as well as the author's exclusive property right on changing the work in question. In order to provide unhindered program usage, the legislator gives authority to a lawful holder to correct errors within the program as well as to perform other necessary changes to the program unless otherwise provided by the contract. Similar purpose has the authority of a legal lawful holder of a computer program copy who is allowed to make a software copy on a permanent computer carrier because a working copy can be damaged or destroyed during its usage. Decompilation which is under aforementioned conditions also covered by the suspension implies the transformation of the so called object code into a source code, [5] which enables the software to function on a certain computer equipment or along with another program, if needed. The only condition is that the data obtained in this way are not allowed to be conveyed to other parties or used for the purposes other than those provided by the law. (Art. 47 parag. 2 LCRR)

Given the exceptional economic importance of the computer program, LCRR provides specific provisions in contract law as well, regarding copyright works created in employment relation which may be waived only if provided by the contract. If the computer program is the object of the contact order, a purcheser acquires the publishing right and all property rights which is in accordance with the conclusion of the contract. When creating other works of autorship, a purcheser, apart from the publishing right, is also given the right to place these works on the market. (Art. 95 LCRR) Finally, it is necessary to emphasize, that an employer is a permanent holder of all property authorities regarding the computer program that is created in the course of the employment relation unless otherwise provided by the contract.

4. Patent Protection of the Computer Program

By accelerated development of computer science and by taking into account the economic value of computer programs, it has been created such legal-political climate which, in some countries, made acceptable that certain types of intellectual creativity related to the software are protected by patents, instead of copyright. [15] Thus, for

example, United States Trademark and Patent Office approved the so-called patents for software back in the late 20th century. In theory, it is stated that the Freeman-Walter-Abele test for evaluation of patentability of mathematical algorithms which the software is based on, has been first adopted in the US.[2] The Examination Guidelines for Computer-Implemented Inventions United States Patent and Trademark Office, has been applied since 1996. [16]

Neither the judicial practice of the member states of the EU, nor the practice of the European Patent Office (EPO) ignore current trends of the American precedents in their own decisions,[17] which is also evidenced by the data showing that thousands of patents are assigned to the computer-applicable inventions in the national authorities and in the EPO. [18] RS has been the member of the EPO from 1st October 2010. In order to join EPC, RS was called upon the decision of the Administrative Council of EPO from 19th June 2008.

Namely, by interpreting the relevant provisions of the European Patent Convention (EPC) [19], the EPO has come to the point of view where the computer program, when it is starting or storing in the computer, produces or has the possibility to produce further technical effect which overcomes normal physical interaction between the software and the computer (hardware) which starts the program, can be patentable.

Similar determination of patentability also exists in the legal system of the US, because the patent can be assigned to the procedure, device or product which is new and useful. Patent protection can also be recognized for the improvement of the procedure, device or product, under the condition that the improvement fulfills the predicted requirements. (United States Code (1952): Title 35 – Patents, Art. 101) In that way, judicial practice in the US deletes the difference between computer programs, as author's works and patentable inventions even more. In the case Allapat, the court took a stand that patent protection can be recognized to the computer program whenever the conditions of patentability are fulfilled. [20] Still, in the legal system of the US, apart from more frequent practice of patenting the computer program, the primary form of protection is the copyright. In the court decisions in the US, during 1960s and '70s, it is distinguished that the source code represents a form of expressing the idea and because of that, the computer program is considered as an author's work. [21] This attitude of the American judicial practice is first regulated by the amendments on the US Copyright Act. An Act for the General Revision of the Copyright Law, Title 17 of the United States Code, and for other purposes, 94th Congress, January 1, 1978. Pub. L. 94−553 (Oct. 19, 1976), and later Computer Software Copyright Act of 1980, Pub. L. No. 96−517, 94 Stat. 3015, 3028 (1980). However, the primacy of the copyright concept, which is widely accepted, does not come into question. Thus, the copyright protects the work's form, i.e. the literary code of the computer program, but not the idea which stands in its background.

In Japan, Patent Office has set special standards for patenting these inventions. According to these standards, software oriented invention can be the object of patent protection if it represents high improved technical idea aimed at using natural laws. This general condition of patentability regarding a software invention is applied when it offers a solution for controling hardware resources or during the procedure related to controling hardware resources, processing information based on phisical or technical characteristics of the object or processing information in which hardware resources are used.[22] Otherwise, in Japanese legal system a copyright concept for the protection of

computer programs was set provided by Japanese Copyright Act (JSA). Computer software was given the status of a special type of authorship works, (not literary ones). However, program languages, rules and algorythms used in computer's program creation are left without copyright protection. (Art. 10 (3) JSA) Therefore, a structure, sequence, organization and program interface, unlike in US legal system, are not covered by the protection. A space in Japanese patent legislation is still being searched to enable the protection of these issues.

EPC and the EPO practice are especially important for every Serbian law in this field. Regarding the EPC, and in Serbian legal system, computer programs are not considered as inventions, unless they have a technical character. Actually, assumption is that computer program *a priori* does not possess a technical character, but if a concrete program has a technical character and fulfills other conditions as well, then it can be considered as an invention. This means that a computer program with technical character can be protected by patent. In the Republic of Serbia, several software patents are approved per year, mostly by domestic inventors. [23]

Conclusion

The copyright protection of computer program in the Serbian law is in accordance with the regulations of the Directive 91/250/EEC, which represents a primary act regulating the issue of the computer program's protection in the EU and it is used by the member states to protect computer programs with copyright, the same as literary works, in terms of the Bern Convention (1971). However, taking into consideration specific characteristics of the computer program as an intellectual property, a consistent application of provisions on literary works copyright is not justified. For this reason, our law as well as certain international regulations, predict a range of provisions by which a need for specific regulation of the computer program is recognized and respected.

On the other hand, the field of computer program application in the increasingly enhanced information era includes even the fields traditionally protected by industrial property rights or patents. The reasons related to its protection originate from this fact. It can be concluded from the analysis of the current judical practice that computer programs, under certain conditions, can be the object of patent protection. The basic requirement considering its patentability is that the object gives technical contribution. However, primary quality of the generally accepted copyright concept does not come into question. KEP and EPO are of special significance for Serbian practice in this field.

Considering that in practice parallel copyright and patent protection of the computer program exist, their relation is also important. We can conclude that these two forms of protection can complement each other so that the copyright is protected by the form in which the computer program is expressed, and the patent ideas on which it is based if they meet the requirements of patentability.

References

1. Smith, B.L.: Balancing Copyright and Patent Protection for Software in the U.s with Consideration for Marketing in Europe, Stockholm, (1998).

2. Šarboh, S.: Zaštita softvera patentom u SAD i u Evropskoj uniji, Pravni život. No. 11, 846, (2004).
3. Berne Convention for the Protection of Literary and Artistic Works.[Online]. Available: http://www.wipo.int/treaties/en/ip/berne/trtdocs_wo001.html.
4. Agreement on Trade-Related Aspects of Intellectual Property Rights. [Online]. Available: www.wipo.int/treaties/en/agreement/trips.html. (current September 2017)
5. Marković, S.: Zaštita računarskih programa patentom, autorskim pravom i pravom *sui generis*, Ljubljana, 12, (1989).
6. Besarovic, V.: Intelektualna svojina, Beograd, 66, (2011).
7. Council Directive 91/250/EEC of 14 May 1991 on the legal protection of computer programs, Official Journal of the European Union, L 122, 17. May 1991, 42-46.
8. Sherwood, R. M.; The TRIPS Agreement: Implications for Developing Countries, Intellectual Property Law Review,Vol. 37, No. 3, 491-543, (1997).
9. Law on the Ratification of the Patent Law Treaty ("Official Gazette of the Republic of Serbia - International Agreements", No. 19, (2010).
10. Marković, S.: Patentno pravo, Beograd, 72 (1997).
11. Davidson, D. M., Davidson, J. A.: Advanced Legal Strategies for Buying and Selling Computers and Software, New York, Chichester, Brisbon, Toronto, Singapore, 13, (1986).
12. Standard Glossary of Software Engineering Terminology, Inc. New York, (1983).
13. Tasić B, Bauer I.: Rečnik kompjuterskih termina, Mikro knjiga, Beograd, (1998).
14. Hoppen, B. N: Software Innovations and Patents - A Simulation Approach, Verlag, Stuttgart, 152 (2005).
15. Idris, K.: Intelektualna svojina, moćno sredstvo ekonomskog rasta, Beograd (2003).
16. Examination Guidelines for Computer-Implemented Inventions“, U.S. Patent and Trademark Office, 1996. Available: www.uspto.gov/web/offices/pac/dapp/pdf/ciig.pdf
17. Ljubojev, N, Sinđelić, S, Legal Protection for Software in United States and Serbian Law, Proceedings, 3rd Scientific-profesional Conference “Textile Science & Economy III”, University of Novi Sad, Technical Faculty “Mihajlo Pupin”, Zrenjanin, Serbia, 10−11, 3(1), 209−216, (2011).
18. Toeniskoetter, S., Protection of Software Intellectual Property in Europe: An Alternative *Sui Generis* Approach, Intellectual Property Law Bultein Vol. 10/2005−2006, 65−81.
[Online]. Available: www.uspto.gov/web/offices/dcom/olia/aipa/PatLaws121. (current Auguste 2017)
19. The Law on Ratification of the European Patent Convention, “Official Gazette SRJ, No. 3, (1996), Annex- International Agreements".
20. Brown, F. V.: The Incompatibility of Copyrigth and Computer Software: An Economic Evoluation and a Proposal for a Marketplace Solution, North Carolina Law Review Vol. 66, No. 981 (1998).
21. Yang, B. G. C.: The Continuing Debate of Software Patents and the Open Source Movement, Texas Intellectual Property Law Jouranl Vol. 13, 175 (2004−2005).
22. Ruping B. K.: Patent Protection of Computer Softwere in Japan and the United States, Institucion of Intellectual Property, Tokyo, 21 (1998).
23. Šarboh, S, Perić, J, Perić, M.: Zaštita softvera - između autorskog prava i patenta, Telekomunikacioni forum, Telfor 2010, Beograd (2010).

Problem solving in software requirements elicitation and specification: Preliminary results from a qualitative study

Tamara Zorić[1], Zeljko Stojanov[1]

[1]University of Novi Sad, Technical faculty "Mihajlo Pupin" Zrenjanin
tamara.zoric@tfzr.rs, zeljko.stojanov@uns.ac.rs

Abstract. Software requirements engineering is highly complex area of software engineering with great influence on overall performance of software projects. All activities in requirements process are highly people dependent, which emphasizes the role of people skills in processing software requirements. Due to natural complexity of requirements processing, problems regularly occur. This study aims at identifying problem solving experience of software engineers in industrial practice. The study uses qualitative research methods for developing thematic descriptions of problem solving practice in elicitation and specification of software requirements. Since this is an ongoing project, the paper presents only preliminary research findings, together with research implications of the study and some further research directions.

Keywords: software requirements, requirements elicitation, requirements specification, problem solving, thematic framework, qualitative study.

1. Introduction

Software requirements describe software systems and reflect needs of software user. Software Requirements Engineering (SRE), or simply Requirements Engineering (RE), is a discipline within software engineering concerned with elicitation, analysis, specification and validation of software requirements. RE assumes joint work of software engineers or business analysts from software organization and software users on determining requirements for software system. Despite recognized importance of software requirements for success of software projects, the practice of acquiring requirements is still far from satisfactory [1]. According to Sommerville [2], the main source of problems in RE is the need to have requirements' descriptions at different levels suitable for different stakeholders. In this context, there exist *user requirements* expressed in natural language and diagrams, and more detailed and technically rich *system requirements* that enable construction of software systems.

Requirements engineering is by nature broad, interdisciplinary, and open-ended, which raises specific groups of problems, such as [3]: (1) investigating goals, functions

and constraints for software system, (2) specifying software system behavior, and (3) managing evolution of software system (including evolution of requirements). Elicitation of suitable requirements is essential for preventing errors, reducing costs and risks and improving quality in software projects [4]. Problems with poor requirements, not identified and solved, usually result with increased costs during software testing (50 times) and maintenance (200 times) [5]. Beecham et al. [6] investigated problems related to software process improvement in 12 software companies. The problems were classified into three larger groups: organizational issues, project issues, and software life cycle issues. In software life cycle group (requirements, design, coding, testing, maintenance), almost a half of all problems are in the field of RE, which strongly emphasizes the importance of identifying and handling RE problems.

By having previous observations in mind, the authors conducted an empirical study aimed at investigating the role of human factor in RE. This paper presents a segment of the study related to problem solving experience in requirements elicitation and specification. In the study participated software practitioners from industry. Since the data analysis has not yet been completed, the paper presents only preliminary results.

2. Related work

Due to its importance for success of software projects, RE has gained attention by research community, which resulted with studies addressing specific aspects of RE. The importance of RE for software projects can be seen from the results of Chaos Report published by The Standish Group [7], in which incomplete requirements, the lack of user involvement, and changing requirements are cited as very critical problems for project success. The focus of this related work section is on typical studies dealing with different types of problems in RE.

Verner et al. [8] organized a survey in Australia and USA aimed at investigating relationships between RE practice and software projects success for in-house requirements management practices. The study was based on a questionnaire that was focused on requirements practice, relationships with customers and users, and project management. The study reported the following problems and directions for improvements: (1) RE methodology must fit into selected software development methodology, (2) requirement should be completed at some stage during project, (3) central requirements repository increase chances for project success, (4) problem domain boundaries should be clearly defined, while requirements could evolve within defined boundaries, and (5) poor requirements leads to wrong estimations, schedules and inadequate staffing within projects.

Solemon et al. [1] presented a study on RE practice and problems in software development companies in Malaysia. The objective of the study is to identify areas for further research and possible improvement directions of the practice. Results suggest that 60% of RE problems originate in RE process. About 50% of problems are consequence of poor communication with customers and users, poor communication among developers, and time and resources allocation. The authors concluded that RE problem management should consider both human and technical problems, with the focus on RE process improvement models.

Fernández et al. [9] reported qualitative analysis of empirical data collected from 228 companies aimed at discovering problems in requirements practice. The authors derived 21 RE problems from the literature, and asked practitioners to rank the five most critical ones. The top cited RE problems regarding criticality for software projects are presented in Table 1. Further, the authors conducted analysis of problem causes and effects for specific contexts, which enables relating causes to problems and identification of RE success factors.

Table 1. Top 10 RE problems (adapted from [9])

RE problem	Total	Cause for project failure
Incomplete and/or hidden requirements	109 (48 %)	43
Communication flaws between project team and the customer	93 (41 %)	45
Moving targets (changing goals, business processes and/or requirements)	76 (33 %)	39
Underspecified requirements that are too abstract	76 (33 %)	28
Time boxing/Not enough time in general	72 (32 %)	24
Communication flaws within the project team	62 (27 %)	25
Stakeholders with difficulties in separating requirements from known solution designs	56 (25 %)	10
Insufficient support by customer	45 (20 %)	24
Inconsistent requirements	44 (19 %)	15
Weak access to customer needs and/or business Information	42 (18 %)	16

3. Case study

An empirical qualitative study was organized with software engineers from industry. The study objective was to investigate the role of human factor in RE, with the focus on communication and cognitive skills. Problem solving is recognized as important aspect of requirements engineering practice tightly related to cognitive skills.

The presented study is only a segment of the whole empirical study that is still underway. This means that data analysis has not yet been completed. Therefore, this paper presents only preliminary or provisional results. These results are very close to the final ones, which the authors expect to get after completing data analysis.

3.1. Context

The study included 14 participants that work as software engineers in industrial practice. The participants were approached based on the authors' contacts.

The average working experience of the participants in software industry is over 10 years (10.29), while the most experienced participant has 30 years of experience. 5 participants work in indigenous software companies, 6 of them work for foreign companies, while 3 of them work as software freelancers.

3.2. Methods

The study was based on qualitative research methods that are regularly used for inquiring human related issues in variety of fields. Qualitative research methods have gained significant attention by software engineering research community since they enable deeper understanding of human related issues of the practice [10, 11].

Data were collected by using semi structured interviews with open ended questions [12]. The interviews were organized in Serbian, while the authors translated the participant answers and research findings to English. The main issue in data analysis and translation was to preserve the meaning of the participants' answers [13]. Field notes were recorded immediately after each interview [14].

Data analysis was based on inductive thematic data analysis technique proposed by Braun and Clarke [15]. Iterative data analysis process is presented in Fig. 1. During the all stages in data analysis, it is common practice to return to the previous stages, and based on the current results to reinvestigate the results of the previous stages. This means, that during development of themes based on initial codes, it is usual to go back to interview transcripts to check how accurately identified theme represents the meaning of participants words (experience).

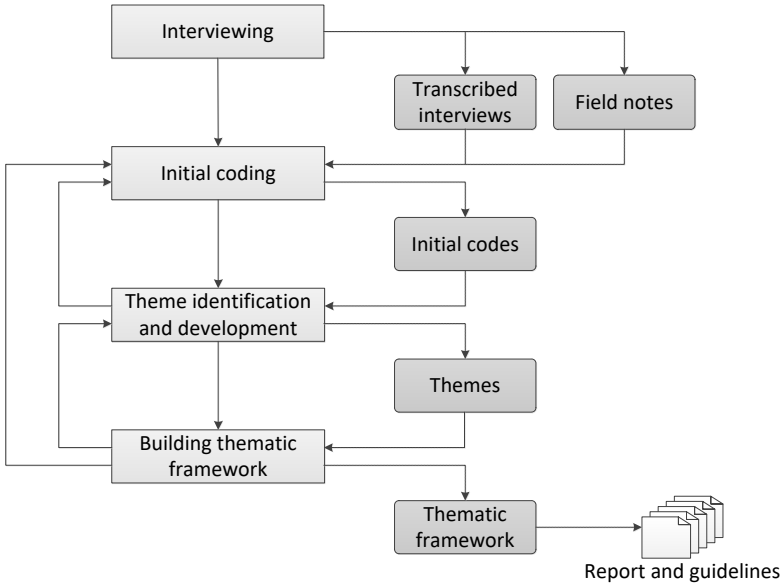


Fig. 1. Process of inductive thematic data analysis used in the study.

All writings related to thoughts, dilemmas and items produced during data analysis were recorded in memos [16]. The descriptions of all initial codes and identified themes were recorded in theoretical memos used for developing themes and thematic framework, while all considerations regarding the context, participants, methods and dilemmas were recorded in methodological memos.

3.3. Findings

Since this study presents only preliminary results, here are outlined only identified themes related to problem solving experience, as well as the typical problem solving process. Final themes and thematic framework are not presented, nor the guidelines that follow from the thematic framework.

Problem solving themes

Regardless of the fact that the research is ongoing, the current findings enable identification of problem-solving **themes**. It was identified that the way in which problems can be solved may depend on the previous experience of software engineers. Majority of the participants believe that any problem can be solved in less time and with less effort if software engineer has experience in solving similar problems.

Since problems can be observed from the aspect of experience, it is clear that problem-solving methods can be observed in the same way. Considering that, Table 2 shows the methods for solving problems in relation to previous experience, derived through data analysis.

Table 2. Methods for solving problems in relation to previous experience

Problem solving methods	Existence of experience	Lack of experience
Using consultancies		X
Trial and error		X
Finding an analogy with previous cases	X	
Using intuition	X	X
Using pen and paper	X	X

The first identified method, whose use clearly requires previous experience, is the method based on finding an analogy with previous cases. The method is often used to solve problems, and it is considered to be a method for rapid problem solving, but also for problem avoiding. When a problem arises, this method can be used by first examining whether there has been any analogous problem. As for the problem avoiding, awareness of the consequences of the problem encourages software engineers to avoid that problem. Bad experience can serve as a warning to avoid making the same decisions.

Given the wide range of problems that can arise in RE field, it is clear that software engineers cannot have experience in solving all types of problems. It has been identified that there are problem solving methods characteristic only for such type of problems. The following two methods have been identified as methods that are characteristic only for solving problems without experience: method of trial and error, and method of solving problems by using consultancy. As for the method of trial and error, software engineers have different opinions on the application of that method. Some software engineers believe that this method can give good results, but in the case of more complicated problems, it is necessary to have more time for trials and repetitions. On the other hand, some software engineers believe that this method does not provide any benefit, and that it is more effective to learn from other peoples' mistakes or to solve

problems with the help of others. In this case, the importance of another mentioned method (method of solving problems by using consultancy) is highlighted, while the method of trial and error is considered the least effective. However, some software engineers believe that this method should be applied every time when there is a lack of experience, or if it is not possible to predict the outcomes and consequences of some solutions. In those cases, this method is considered to be the best way to discover and define how system works.

As for consulting as a way of problem-solving, it is considered that consultation can compensate the lack of experience and knowledge in a particular field. Consultation includes consultation with experts and colleagues, consultation of literature and information retrieval by researching on the Internet. The aim of this method is to provide a better understanding of the problem.

In addition to these methods, it has been identified that there are methods whose use does not depend on previous experience. These methods are related to solving problems intuitively, as well as solving problems by using pen and paper. Using pen and paper makes it easier to solve problems, because it allows the problem to be visualized.

Problem solving process

Participants in the research believe that the problems need to be solved at the moment they occur, without delay. They emphasized the importance of quickly solving problems and pointed out that there is a process for it.

The problem solving process depends on the problem itself and the method chosen for solving it. Basically, the process should begin by examining the cause of the problem, and then continue by selecting a method for solving it, and finally applying that method. The findings of the research also point out to certain activities that can be involved in a problem solving process. Fig. 2 presents the most general problem solving process in RE.

The findings of the research indicate that the first step in problem-solving process is to identify and examine the cause of the problem. The cause of the problem can be examined by asking additional questions in order to get acquainted with the problem. The participants in the research pointed out that, when the problem is identified, a software engineer should not promise a solution to this problem immediately unless he is certain that it can be achieved.

The next step in this process is to divide the problem into smaller parts. The participants pointed out that the division of a problem into smaller segments is one of the basic principles of resolving problems, regardless of the problem domain. This step is considered important due to the fact that by dividing the problem it is possible to see what the problem involves, and to solve each part separately. The whole work is divided into tasks, and tasks are assigned to each part in order to solve the problem as a whole.

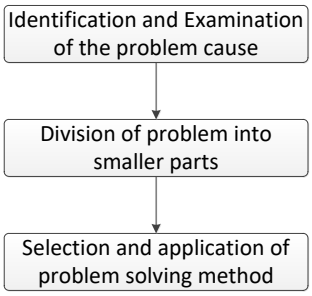


Fig. 2. The most general problem solving process.

The last step is to select the method with which the problem will be solved. As it is previously explained, the method can be selected according to whether the necessary previous experience exists. Based on the selected method, the process can be resumed differently. In case the method of finding an analogy with previous cases is selected, solving the problem will immediately follow. On the other hand, the method of trial and error will require iterations in the problem solving process. Iterations include problem solving attempts, on the basis of which errors can be identified. After each attempt, feedback needs to be obtained to determine if there are any errors.

3.4. Validity

Since this paper present preliminary results, specific validity issues should be identified and discussed in order to increase credibility of the presented findings and the study. Descriptive, interpretative and theoretical validity as defined by Maxwell [17], related to development of theoretical constructs from empirical data, were ensured through careful implementation of all steps in inductive thematic analysis [15] and extensive reflexive writing in memos during all phases of the research [18]. However, the authors are aware that comprehensive discussion of descriptive, interpretative and theoretical validity is only possible after completing the analysis. Additional steps to ensure the validity of the developed themes is to use member check technique to discuss developed constructs with some of the participants [19], which will be done after completing the thematic framework. External validity, referring to generalizability or transferability of the research findings [17], is always problematic in qualitative research, since the whole research design and findings are usually highly context dependent. However, the authors expect that final results will provide useful lessons for experts from software industry and education.

4. Implications

Although this paper presents only preliminary results related to a segment from empirical qualitative study, several implications and benefits can be drawn for different people interested in RE practice.

First of all, practitioners from industry can find explanations of some typical problems from the practice and links to relevant literature, as well as some preliminary guidelines for staffing a team based on required problem solving skills (if they properly interpret the results).

Researcher can find brief guidelines for organizing qualitative study with experts from the field, as well as they can use identified themes for deeper inquiry.

Educators can find themes related to problem solving identified by the practitioners from software industry, and based on that include suitable topics in university courses.

5. Conclusions

Due to high complexity of RE process and dependency on human factor, problem regularly occur in RE practice, which require software engineers with different problem solving experience and skills. Presented study and findings contribute to understanding problem solving experience of software engineers in practice. Identified themes and typical problem solving process can be used as staring points in further research, as well as guidelines for introducing practice oriented topics in software engineering courses at university.

Further work firstly assumes finishing data analysis and providing a thematic framework with comprehensive view of problem solving experience in elicitation and specification of software requirements. Based on the theoretical framework, several guidelines can be proposed for improving RE practice, staffing in software organizations, or including relevant themes in software engineering education.

Acknowledgement

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, 2011-2017.

References

1. Solemon, B., Sahibuddin, S., Ghani, A. A. A.: Requirements Engineering Problems and Practices in Software Companies: An Industrial Survey. In: Proceedings of International Conference on Advanced Software Engineering and Its Applications, ASEA 2009: Advances in Software Engineering, pp 70-77. Springer, Berlin, Heidelberg. (2009).
2. Sommerville, I.: Software Engineering, 9th edition. Addison-Wesley, Boston, MA, USA. (2011).
3. Zave, P.: Classification of research efforts in requirements engineering. ACM Computing Surveys, Vol. 29, No. 4, 315-321. (1997).

4. Procaccino, J. D., Verner, J. M., Overmyer, S. P., Darter, M. E.: Case study: factors for early prediction of software development success, Information and Software Technology, Vol. 44, No. 1, 53-62. (2002).
5. Leffingwell, D., Widrig, D.: Managing Software Requirements: A Unified Approach, Addison-Wesley, Reading, MA, USA. (2000).
6. Beecham, S., Hall, T., Rainer, A.: Software Process Improvement Problems in Twelve Software Companies: An Empirical Analysis. Empirical Software Engineering, Vol. 8, No. 1, 7-42. (2003).
7. The Standish Group. Chaos Report. 2014.
8. Verner, J., Cox, K., Bleistein, S., Cerpa, N.: Requirements Engineering and Software Project Success: an industrial survey in Australia and the U.S. Australasian Journal of Information Systems, Vol. 13, No. 1, 225-238. (2005).
9. Fernández, D.M., Wagner, S., Kalinowski, M. et al.: Naming the pain in requirements engineering: Contemporary problems, causes, and effects in practice. Empirical Software Engineering, Vol. 22, No. 5, 2298–2338. (2017).
10. Dittrich, Y., John, M., Singer, J., Tessem, B.: For the Special issue on Qualitative Software Engineering Research. Information and Software Technology, Vol. 49, No. 6, 531-539. (2007)
11. Dybå, T., Prikladnicki, R., Rönkkö, K., Seaman, C., Sillito, J.: Qualitative research in software engineering. Empirical Software Engineering, Vol. 16, No. 4, 425–429. (2011).
12. King, N., Horrocks, C.: Interviews in Qualitative Research. SAGE Publications. London, UK. (2010).
13. Stone, C., West, D.: Translation, representation and the Deaf 'voice'. Qualitative Research, Vol. 12, No. 6, 645-665. (2012).
14. Wolfinger, N. H.: On writing fieldnotes: collection strategies and background expectancies, Qualitative Research, Vol. 2, No. 1, 85-93. (2002).
15. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qualitative Research in Psychology, Vol. 3, No. 2, 77-101. (2006).
16. Birks, M., Chapman, Y., Francis. K.: Memoing in qualitative research: Probing data and processes. Journal of Research in Nursing, Vol. 13, |No. 1, 68-75. (2008).
17. Maxwell, J. A.: Understanding and Validity in Qualitative Research, Harvard Educational Review, Vol. 62, No. 3, 279-300. (1992).
18. Watt, D.: On Becoming a Qualitative Researcher: The Value of Reflexivity. The Qualitative Report, Vol. 12, No. 1, 82-101. (2007).
19. Buchbinder, E.: Beyond Checking: Experiences of the Validation Interview. Qualitative Social Work, Vol. 10, No. 1, 106–122. (2011).

Machine Learning in Transportation and Logistics Systems

Vladimir Brtka[1], Gordana Jotanović[2], Eleonora Brtka[1], Ivana Berković[1], Višnja Ognjenović[1]

[1] University of Novi Sad, Technical faculty "Mihajlo Pupin"
23000 Zrenjanin, Serbia
{vbrtka, norab, berkovic, visnjao}@tfzr.uns.ac.rs
[2] Faculty of Transport and Traffic Engineering, University of East Sarajevo
Doboj, Bosnia and Herzegovina
gjotanovic@yahoo.com

Abstract. The paper deals with the business process of freight forwarding company which is structured into incoming and outgoing transport legs jointly aim at ensuring that freight is timely delivered to customers. Each leg may contain up to four segments or hops which are additional airline routes. The main goal of this research is to investigate the impact of incoming transport legs to the number of hops of the outgoing transport leg. The analysis was done by tools based on Rough sets theory. Two parameters are singled out: Actual duration of first incoming leg and actual duration of outgoing transport leg.

Keywords: machine learning, rough sets theory, freight forwarding

1. Introduction

Transportation and logistics systems are often in the form of networks and are among the most complex artifacts developed by humans [1]. In order to understand these complex systems, models are produced. Numerous parameters of models are tweaked until the predictions of the model agreed with empirically collected evidence. But, usually it is hard to collect empirical evidence although sensors offer a novel approach to data collection. The problem to be resolved is to find minimal expected-time path through a transportation route, usually a road network. This is environment-navigation problem that can be decomposed to multiple sub-tasks, but environment has to be presented by a model and a decision making procedure must be adopted. In [1] the k-robust algorithm is proposed for environments with high volatility while taking advantage of the up-to-minute sensor input. Shortest path finding is often done with simple and well known graph search algorithm in the domain of Artificial Intelligence (AI), named A* (A star). Decision making problem sometimes can be modeled as a Markov decision process [2, 3] or Semi-Markov decision process [4, 5].

While dealing with transportation management issues, Machine Learning (ML) models and algorithms are used, such as [1, 6, 7, 8]: clustering, artificial neural networks, the mixture of Gaussian trees generative model, the sequence label realignment framework, dimensionality reduction algorithms such as principal component analysis, decision trees, regression models, etc.

The business process is structured into incoming and outgoing transport legs, which jointly aim at ensuring that freight is timely delivered to customers. This research investigates the impact of incoming transport legs characteristics to the number of hops of the outgoing leg. The investigation is done by the application of the rough sets theory.

2. Materials and Methods

The source of data used in this investigation is a case study of the business processes of a freight forwarding company, in which up to three smaller shipments from suppliers are consolidated and then shipped together to customers [9, 10, 11]. Data collection and preparation were done by Transport and Logistics Case Study Data Set (Cargo 2000) researches, so no actual data were collected in this particular research. There are incoming and outgoing transportation legs which converge in order at ensuring that freight is timely delivered to customers. Methods used are based on the application of the Rough sets theory.

2.1. Materials

The data originates from a forwarding company's Cargo 2000 system for a period of five months and includes traces of 3,942 actual business process instances, comprising 7,932 transport legs and 56,082 service invocations. Each execution trace includes planned and effective durations in minutes for each of the services of the business process. The data were previously subjected to the process of anonymization, while undefined instances and instances with missing values were removed for the purpose of this particular investigation.

According to Cargo 2000 (Cargo iQ) industry standard transport services are denoted by three-letter acronyms. This is an initiative of the International Air Transport Association (IATA). Data collection for each of the transport legs involves the following physical transport services: RCS: Check in freight at departure airline, DEP: Confirm goods on board, RCF: Accept freight at arrival airline, and DLV: Deliver freight. As a transport leg may involve multiple segments activity RCF may loop back to DEP, while the number of segments per leg ranges from one to four.

Additional data pre-processing task were done in this investigation: all unique attributes were excluded; also phases DEP and RCS were excluded. These are repeating phases, defined by number of hops in the leg, so that number of hops describes the repetition count of DEP and RCS in each leg.

2.2. Data Preparation

There were 3942 records in the original dataset, after omitting incomplete records, there were 2624 records left. Remaining attributes are described in Table 1.

Table 1. Attributes after dataset pre-procesing and their descriptions.

Num.	Attribute	Description
1	i1_dlv_p	Planed duration of incoming leg 1
2	i1_dlv_e	Actual duration of incoming leg 1
3	i1_hops	Number of segments (hops) in leg 1
4	i2_dlv_p	Planed duration of incoming leg 2
5	i2_dlv_e	Actual duration of incoming leg 2
6	i2_hops	Number of segments (hops) in leg 2
7	i3_dlv_p	Planed duration of incoming leg 3
8	i3_dlv_e	Actual duration of incoming leg 3
9	i3_hops	Number of segments (hops) in leg 3
10	o_dlv_p	Planed duration of outgoing transport leg
11	o_dlv_e	Actual duration of outgoing transport leg
12	o_hops	Number of segments in the outgoing transport leg

Transport leg may involve multiple segments (hops), meaning that incoming leg may be realized via additional airline routes (up to four in this case). However, planed duration, as well as actual duration in minutes of departure segment (DEP) and arrival segment (RCF) were excluded. These values were usually undefined because most of the legs had no additional segments. Still, transportation route does not have to contain all legs, so that there are undefined values in the dataset marked with "?". Outgoing leg may also contain some additional hops.

2.3. Methods

Pawlak [12, 13, 14, 15] introduced Rough Set Theory (RST) in the early 1980s as a way to reason about imprecise and uncertain data. Information system I is defined by a pair (U, A), where U is a non-empty finite set of objects and A is a non-empty finite set of attributes. Set V_a is domain set of attribute a, or so called value set, so that $a(x)$ refers to the value of attribute a for the object $x \in U$. Each object is represented by means of attributes and their values. As a starting point there is a two-dimensional data table or flat table where rows represent objects, while columns represent attributes. Based on the inability to distinguish between objects RST deals with the approximation of sets and is suitable for classification tasks. RST offers a good mechanism for decision rule synthesis. These rules are usually of the If-Then form: $Ant \rightarrow Con$ where Ant is antecedent part of the rule while Con is the consequent part of the rule. An approximating feature of RST enables rule generation so that consequent part contains OR logical operator. In the pre-processing step numerical attributes (attributes with numerical values) may have to be discretized, so that intervals or ranges are used

instead of the exact continuous values. Rule generation is done in a two-stage process: first, minimal attribute subsets are computed (so called reduct sets) and second, rules are generated from these minimal sets of attributes. Information system I generates a discernibility $|U| \times |U|$ matrix M_I (1) where each entry $M_I(x, y) \subseteq I$ consists of the set of attributes that can be used to discern objects $x, y \in U$.

$$M_I(x, y) = \{a \in A: discerns(a, x, y)\} \quad (1)$$

The $discerns$ function is defined as in (2):

$$discerns(a, x, y) \Leftrightarrow \mathrm{a(x)} \neq \mathrm{a(y)} \quad (2)$$

As shown in (3), discernibility matrix generates an indiscernibility relation $Ind_A \subseteq U^2$, this is binary relation that is reflexive, symmetric and transitive.

$$xInd_A y \Leftrightarrow M_I(x, y) = \emptyset \quad (3)$$

The inidscernibility set of an object $x \in U$ denoted by $Ind_A(x)$ is defined by (4):

$$Ind_A(x) = \{y \in U: xInd_A y\} \quad (4)$$

However, not all attributes are needed in order to preserve indiscernibility relation, those attributes that are redundant are called superfluous and can be omitted. Set of attributes A is divided to non-empty set of condition attributes C and non-empty set of decision attributes D so that $C \cup D = A$ and $C \cap D = \emptyset$. Rules exploit functional dependences between condition attributes that are included in the antecedent part of the rule and decision attributes that are included in the consequent part of the rule. Usually, there is only one binary decision attribute. The set is approximated by means of lower and upper approximation. Analogous algorithm for decision rules synthesis is Syntax systematic Classification of Objects (SSCO) [16, 17, 18]. This is an iterative algorithm based on depth first search (DFS) strategy and indiscernibility graphs. The SSCO takes into account attribute priorities and can be interrupted in case of extensive memory/time consumption.

3. Results

After pre-processing steps, the dataset was loaded to the free Rosetta system (A Rough Set Toolkit for Analysis of Data, supported by the Knowledge Systems Group, NTNU, Norway). Attributes 1-11 from Table 1 are condition attributes, while attribute 12 is decision attribute. All in all, 2201 rules were generated, many of them include OR logical operator in the If part of the rule. Table 2 shows five rules with highest support.

Table 2. Rules with highest support

Num.	IF	THEN	Support
1	i1_dlv_p([3765, 3983)) AND i1_dlv_e([*, 367)) AND i1_hops([*, 2)) AND i2_dlv_p([3763, 3990)) AND i2_dlv_e([*, 359)) AND i2_hops([*, 2)) AND i3_dlv_e(?) AND o_dlv_p([3770, 3930)) AND o_dlv_e([2182, *))	o_hops(1)	22(22)
2	i1_dlv_p([3765, 3983)) AND i1_dlv_e([*, 367)) AND i1_hops([*, 2)) AND i2_dlv_p([3763, 3990)) AND i2_dlv_e([*, 359)) AND i2_hops([*, 2)) AND i3_dlv_e(?) AND o_dlv_p([3770, 3930)) AND o_dlv_e([*, 624))	o_hops(1) OR o_hops(2)	17(16,1)
3	i1_dlv_p([3765, 3983)) AND i1_dlv_e([367, 1544)) AND i1_hops([*, 2)) AND i2_dlv_p([3763, 3990)) AND i2_dlv_e([*, 359)) AND i2_hops([*, 2)) AND i3_dlv_e(?) AND o_dlv_p([3770, 3930)) AND o_dlv_e([*, 624))	o_hops(1)	11(11)
4	i1_dlv_p([3765, 3983)) AND i1_dlv_e([*, 367)) AND i1_hops([*, 2)) AND i2_dlv_p([3763, 3990)) AND i2_dlv_e([*, 359)) AND i2_hops([*, 2)) AND i3_dlv_e(?) AND o_dlv_p([3770, 3930)) AND o_dlv_e([624, 2182))	o_hops(1)	11(11)
5	i1_dlv_p([3765, 3983)) AND i1_dlv_e([367, 1544)) AND i1_hops([*, 2)) AND i2_dlv_p([3763, 3990)) AND i2_dlv_e([*, 359)) AND i2_hops([*, 2)) AND i3_dlv_e(?) AND o_dlv_p([3770, 3930)) AND o_dlv_e([624, 2182))	o_hops(1) OR o_hops(2)	10(9,1)

The If parts of the rules were generated from reduct set: {i1_dlv_p, i1_dlv_e, i1_hops, i2_dlv_p, i2_dlv_e, i2_hops, i3_dlv_e, o_dlv_p, o_dlv_e} inferred by application of the variation of Johnson's reduct calculating algorithm. The length of this reduct set is nine, meaning that two condition attributes are superfluous. The generated rules indicate the properties required for the number of outgoing hops to be 1 (rules 1, 3, 4) and the properties required for the number of outgoing hops to be 1 OR 2 (rules 2, 5). Support consists of two values Left hand side support (If part of the rule) and Right hand side support (Then part of the rule). If these two values are different, then OR logical operator is present in the Then part of the rule.

4. Conclusions

The research deals with actual business process of forwarding company including airline transportation. There are incoming transport legs combined to form outgoing transport leg in the timely acceptable manner. Every leg may contain up to four segments or hops meaning that additional airline routes are used. Starting from the data that originates from Cargo 2000 system for a period of five months, the impact of incoming legs characteristics to number of hops in the outgoing leg is investigated. It is concluded that number of hops in the outgoing transport leg depends on: Actual duration of first incoming leg and actual duration of outgoing transport leg (see rules from Table 2) since there are multiple values of these attributes, while the values of other attributes included to the If part of rules are same. Future research will investigate the impact of incoming legs properties to the difference between planed and actual duration of outgoing transport leg.

ACKNOWLEDGMENT

Ministry of Science and Technological Development, Republic of Serbia financially supported this research, under the project number TR32044: "The development of software tools for business process analysis and improvement".

References

1. T.·Šingliar, Machine Learning Solutions for Transportation Networks, PhD thesis, University of Pittsburgh. (2008)
2. D. P. Bertsekas, Dynamic programming and optimal control. Athena Scientific, 1995.
3. D. P. Bertsekas, J. N. Tsitsiklis, An analysis of stochastic shortest path problems, Mathematics of Operations Research, 16(3). (1991)
4. R. A. Howard, Semi-Markovian decision processes, In Proceedings of International Statistical Inst., Ottawa, Canada. (1963)
5. M. L. Puterman, Markov decision processes: discrete stochastic dynamic programming. John Wiley, New York. (1994)
6. E. Alpaydın, Introduction to Machine Learning, MIT Press, ISBN: 978-0-262-01243-0. (2010)
7. N. J. Nilsson, Introduction to Machine Learning, Robotics Laboratory, Department of Computer Science, Stanford University. (1998)
8. A. Smola, S.V.N. Vishwanathan, Introduction to Machine Learning, ISBN: 0-521-82583-0, Cambridge University Press. (2008)
9. A. Metzger, P. Leitner, D. Ivanovic, E. Schmieders, R. Franklin, M. Carro, S. Dustdar, and K. Pohl, " Comparing and combining predictive business process monitoring techniques," IEEE Trans. on Systems Man Cybernetics: Systems. (2015)
10. A. Metzger, R. Franklin, and Y. Engel, " Predictive monitoring of heterogeneous service-oriented business networks: The transport and logistics case," in Service Research and Innovation Institute Global Conference (SRII 2012), ser. Conference Publishing Service (CPS), R. Badinelli, F. Bodendorf, S. Towers, S. Singhal, and M. Gupta, Eds. IEEE Computer Society. (2012)
11. Z. Feldmann, F. Fournier, R. Franklin, and A. Metzger, "Industry article: Proactive event processing in action: A case study on the proactive management of transport processes," in Proceedings of the Seventh ACM International Conference on Distributed Event-Based Systems, DEBS 2013, Arlington, Texas, USA, S. Chakravarthy, S. Urban, P. Pietzuch, E. Rundensteiner, and S. Dietrich, Eds. ACM. (2013)
12. Z. Pawlak, J. Grzymala-Busse, R. Slowinski, W. Ziarko, Rough Sets, Association for Computing Machinery. Communications of the ACM, 38, 11 ABI/INFORM Global, pg. 89, 1995.
13. Z. Pawlak, A. Skowron, Rudiments of rough sets, J. of Information Sciences 177, pp 3–27, 2007.
14. Z. Pawlak, Rough set approach to knowledge-based decision support, European Journal of Operational Research, 99, pp. 48–57, 1997.
15. Z. Pawlak, Rough Classification, Int. J. Human–Computer Studies 51, pp. 369–383, 1999.
16. V. Ognjenovic, V. Brtka M. Jovanovic, E. Brtka, I. Berkovic, The Representation of Indiscernibility Relation by Graph, Proceedings of 9th International Symposium on Intelligent Systems and Informatics, pp. 91-94, 2011.
17. V. Brtka, I. Berkovic, E. Stokic, B. Srdic, Automated extraction of decision rules from medical databases - A rough sets approach, Proceedings of 5th IEEE International Symposium on Intelligent Systems and Informatics, pp 27–31, 2007.
18. V. Brtka, I. Berkovic, E. Brtka, V. Jevtic, A Comparison of Rule Sets Induced by Techniques Based on Rough Set Theory, Proceedings of 6th IEEE International Symposium on Intelligent Systems and Informatics, pp 354–357, 2008.

Blockchain Technology for Health Data Exchange

Dejan Lacmanović[1], Predrag Pecev[1], Branko Markoski[1], Sanja Stanisavljev[1] and Dragica Radosav

[1] Technical Faculty "Mihajlo Pupin"
Djure Djakovica bb,
23000 Zrenjanin, Serbia
dlacman@tfzr.uns.ac.rs, markoni@uns.ac.rs, pecev@tfzr.uns.ac.rs, sanja@tfzr.uns.ac.rs, radosav@tfzr.uns.ac.rs

Abstract. This paper describes what blockchain technology is and how implementation of blockchain technology can improve the security and costs for accessing and updating health care provider data. This paper discussed how the identity mechanism inherent to a blockchain, can blockchain support the medical data exchange, and how cryptography can support the secure authorization of data across a distributed network. Infrastructure technologies will play a major role in supporting the development of these new networks. As the health care industry looks for novel ways to improve interoperability, blockchain technology is an innovation that can powered health information exchange and unlock the true value of interoperability and scalability.

Keywords: blockchain, information systems, health care, infrastructure.

1. Introduction

Blockchain is a technology devised by a group of people known by the pseudonym Satoshi Nakamoto and it was originally designed for Bitcoin – a digital currency. It allows digital information to be distributed not copied this forming a backbone for the new type of internet. Nowadays the community is finding new potentials for usage of this technology [1].

Blockchain technology concepts and a distributed database concepts are quite similar. In a way, data in a blockchain technology can be looked as a spreadsheet that is duplicated many times across the network, and there lies a similarity with a distributed database concept. The key advantage is that a network is designed to regularly update aforementioned spreadsheet across the network. Data in this network is not centralized, it is public and easily verifiable, hosted by a vast number of computers at the same time and thus accessible to anyone on the internet. In a way, Blockchain can be described as a distributed document database that does not have an authoritative master copy, thus forming the decentralized feature.

By storing data in form of a blockchain the data that are identical across the network the blockchain does not have a single point of failure and it cannot be fully controlled by a single entity since there is no "master copy" of a blockchain. Blockchain technology is transparent and incorruptible. Here data is stored in a linked list of record called blocks that are secured using various cryptography methods. Each block contains

a link to a previous block (usually a hash pointer), a timestamp and data. Also, the validity of every piece of data in the blockchain can be independently confirmed by every participant.

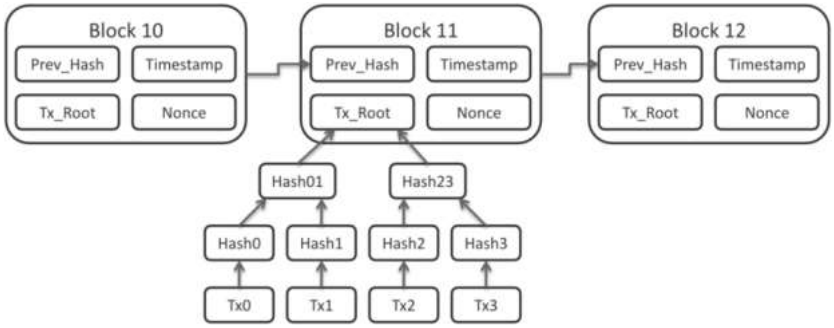


Figure 1. Blockchain structure [1]

Blockchain technology enable internet users to create value and authenticate digital information. Having this in mind Blockchain technology can be applied towards : creating smart contacts, crowd funding, e-Governance, supply chain auditing, file storage, Internet of Things (IoT), data management, identity management, AML and KYC (Anti-money laundering and know your customer), stock trading and many other applications.

Database technology is not new, distributed databases have been around for a decade, and relational databases have existed for even longer. Blockchain is another form of database, and they share many elements with traditional database forms. Differences are that make them truly innovative. Blockchain is intended to be shared, by individuals, organizations, and even devices. In a digital world, the databases are the infrastructure, so a blockchain is also one form of infrastructure. What what's make a blockchain technology so unique? [2]

Identity

Blockchain contain a built-in identity mechanism - a cryptographically secure public-private key pair - used to associate activity on the network with a specific participant (e.g., person, entity, device...). By itself, the key pair is pseudonymous, not revealing the participant's actual identity. However, supplemental information, such as name, contact information, or professional credentials, can be associated with a key pair, merging on-chain and off-chain identities. In a health care context, blockchain' unique identity mechanism could provide the foundation for a unified patient ID across payers and providers [2, 7].

Permission gradient

Using the identity system as a foundation, permissions can be assigned to participants on a network. Permissions correspond with certain abilities on the blockchain, such as the ability to read or write data. Permissions can be enabled to individuals' at the most granular level, for example: an individual could be granted the permission to read and write to Document A, but only the ability to read Document B. Because these permissions are also stored on the blockchain, a participant in the network can be certain that the uploaded data is only accessible by the party to whom access was granted, despite this data being hosted in a decentralized manner [2, 7].

Due to improvements in genetic research and the advancement of precision medicine, health care is witnessing an innovative approach to disease prevention and treatment that incorporates an individual patient's genetic makeup, lifestyle and environment. Simultaneously, IT advancement has produced large databases of health information, provided tools to track health data and engaged individuals more in their own health care. Combining these advancements in health care and information technology would foster transformative change in the field of health IT [3].

2. Blockchain in health care data records

A blockchain powered health information exchange could unlock the value of interoperability. Blockchain-based systems have the potential to increase security and reduce the costs of health care data transactions and records. For these and other high-potential areas, determining the viability of the business case for blockchain is paramount to realize the benefits of improved data integrity, decentralization and disintermediation of trust, and reduced transaction costs. The exchange of Personal Health Records and Health Information Exchange (HIE) data via the Integrating the Health care Enterprise protocol is an important part of addressing the challenges of system interoperability and accessibility of medical records. The strategy outlined to date provides the technical requirements and specific incentives for health systems to meet the interoperability standards necessary to support the envisioned Health Information Network, buttressed by a network of Health Information Exchange data operating on a broad scale. The health systems are focused on private exchanges and are looking for low cost solutions that enable secure integration and support the assembly of virtual health systems that move beyond organizational boundaries [4].

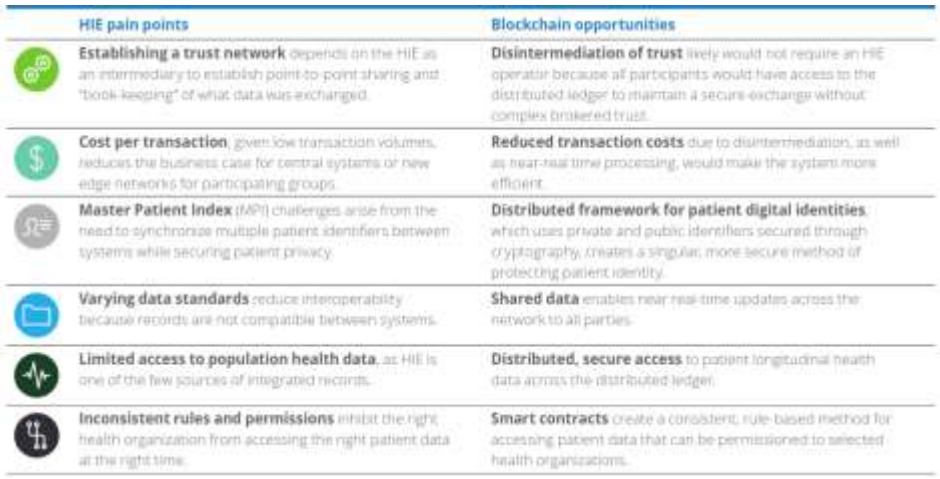


Figure 2. Blockchain opportunities in health information exchange [4]

Blockchain technology is not a magic bullet for data standardization, but offer a promising new distributed framework to amplify and support integration of health care

information across a range of uses. It addresses several existing pain points and enables a system that is more efficient, disintermediated, and secure [4].

3. Blockchain technology limitations

Blockchain are limited by the developing state of the technology and certain design elements inherent to distributed systems. Because a blockchain is technically of shared ledger, each node on the network maintains a copy of the ledger. This means that any data stored on the blockchain is duplicated at each node. Because of that, it would be inefficient and unnecessarily duplicative, to store large amounts of data on a blockchain. Instead, as we suggest in this paper, blockchain technology can be used to connect off-chain data stores, acting as an identity and permissioning infrastructure between parties on the network [2].

Scalability Activity on blockchain networks has increased every year since bitcoin was released in 2009, but the protocols that exist today are not quite ready to support the speed and volume requirements of the health industry. In a blockchain context, speed and security are often inversely related, more of one means less of the other. This is due in part to a blockchain, system's need to have consensus, or agreement on the current state of the ledger. In bitcoin, system's look for consensus as to the state of bitcoin balances across all addresses on the network. In the context of a proposed distributed payer network, system's look for consensus as to the state of payer access permissions to provider data files across the network. In either case, consensus requires some amount of computation and time. The amount of each will decrease over time, but it is a limitation of current blockchain systems and something that must be overcome before a blockchain can replace current production systems [5, 6].

Before blockchain-based applications can be widely adopted, Technology standards must be developed. Several competing protocols exist, as bitcoin, ethereum, hyperledger, etc. There are also a handful and application development suites for each protocol. Meaningful vertical development and scalability of the health care industry will only be achieved once a standard has been established. Impact dependent on middleware and application layers. The internet as we know it is built on foundational protocols such as TCP/IP and DNS (Transmission Control Protocol/the Internet Protocol and domain name protocols, respectively). These protocols support and enable consumer-facing applications such as the Web and email. Like TCP/IP and DNS, blockchain is an infrastructure technology, and its value for health care will only be realized by the middleware and consumer-facing applications it supports. The proposed data management approach in this paper is one example of an application, and there will be countless others, but development takes time [2].

4. Conclusion

Blockchain technology creates unique opportunities to reduce complexity, enable trustless collaboration, and create secure and permanent information. Health Care organizations is right to track this rapidly evolving field to identify trends and sense

areas where government support may be needed for the technology to realize its full potential in health care. To shape blockchain's future, Health Care organizations should consider mapping and assemble the blockchain ecosystem, establishing a blockchain framework to coordinate early developing systems and supporting a discovery of models for health care data records exchange. Strongest point of blockchain technology is potential increases with the number of participants in the network, but to derive value from the network for all participants, a common approach is needed. For health care organizations, it is necessary to evaluate which information should be stored on or off the blockchain and the format in which it should be stored. Blockchain technology, presents numerous opportunities. A blockchain-enabled, trusted exchange of health information can provide longitudinal views of patients' health, generate new insights about population health, and support the move toward value-based care. With greater transparency, trust, and access to data, health care organizations can then also garner insights for better safety, effectiveness, quality, and security of foods, drugs, vaccines, and medical devices. The promise of blockchain has widespread implications for stakeholders in the health care ecosystem. In the long term, a nationwide blockchain network may improve efficiencies and support better health outcomes for patients.

ACKNOWLEDGMENT

Provincial secretariat for higher education and scientific research, Republic of Serbia, Autonomous Province of Vojvodina financially supported this research, under the project application number 142-451-3823/2016-01: "Application of Blockchain technology to effectively reduce obesity in school-age children".

References

1. Meshed Insights Ltd, What Is A Blockchain? 2015. [Online]. Available: https://meshedinsights.com/2015/12/18/what-is-a-blockchain/
2. The Linux Foundation, Blockchain in health, 2016. [Online]. Available: www. hyperledger.org/
3 Beck, R., Stenum Czespluch, J., Lollike, N. & Malone, S. Blockchain – the Gateway To Trust-Free Cryptographic Transactions. In ECIS 2016 Proceedings, 2016.
4. Krawiec RJ, Housman D, White M., Filipova M, Quarre F., Barr D., Nesbitt A., Fedosova K., Killmeyer J, Israel A, Tsai L, Blockchain: Opportunities for Health Care, Deloitte, 2016.
5. Salmony, M.. Blockchain – not for Payments? BIT – Banking and Information Technology, 56(2), pp.6–8, 2016.
6. Tapscott, D. & Tapscott, A., Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business and the World, New York: Penguin, 2016..
7. Swan, M., Blockchain: Blueprint for a New Economy, Sebastopol: O'Reilly, 2015.

Frameworks and Design Patterns in Software Development: A Review

Amar Kansara[1], Ljubica Kazi[2], Zoltan Kazi[2] and Dejan Radovanovic[2]

[1] Parth Systems, 3-1293 Zaveri Road, Navsari 396 445, Gujarat, India
[2] University of Novi Sad, Technical faculty "Mihajlo Pupin" Zrenjanin, Djure Djakovica bb, 23000 Zrenjanin, Serbia
amar.kansara@gmail.com, ljubica.kazi@gmail.com, zoltan.kazi@gmail.com, radoovanovic@yahoo.com

Abstract. Contemporary software development has evolved from crafting to industry with standard methodologies, components and working environments. Software industry recognize teamwork collaboration, development agility and code quality as key success features that leads to disciplined approach supported by software design patterns and frameworks. They enable general and abstract software components for certain software categories. This way, by using design patterns and frameworks, the code could be more readable within teamwork, it could be produced faster and could be more reliable (since it is composed of components that were already tested and used for similar software). Aim of this paper is to represent analysis results of related research in domain of software architectures and application of software design patterns and frameworks. Also, some of the most popular design patterns and frameworks will be described.

Keywords: software architecture, software composition, framework, design pattern, meta-models.

1. Introduction

Building reusable software components is an important task in software engineering today, because it can improve the productivity and quality of software. The basic idea in building reusable software components is to design and build interchangeable software parts that reduce the time and energy in developing software. Some of the experiences from multiple companies studies (such as Nippon, Toshiba, DEC, Hewlet-Packard, NASA and others) reported quality improvement, defect reduction, increase of productivity with reduction in time to market, reduction in customer complaints. Each of indicators had average success of 50-75%, which lead to conclusion of return of investment (ROI) in introducing reusability of software [1].

Software developers typically build applications using frameworks. Since it is costly and time-consuming to build software, frameworks enable applications to be assembled by integrating pluggable components. In addition, frameworks isolate applications development from a variety of complex tasks. [6] This is a facility that software developers use to spend more time in developing the software support to the requirements, rather than preparing the tools for application development. Programmers can build applications quickly with the frameworks, through the sharing of common classes, collections of functions and procedures. [1] During the

software building, the successful solution for overcoming software complexity is in using design patterns that capture the static and dynamic structures and collaborations among components. [7] [8]

Aim of this paper is to analyze related research in the domain of software architecture and composition, frameworks and design patterns and their using for general software development and for the software of specific purpose and categories, such as web applications, mobile applications, multimedia etc.

2. Theoretical Background

Basic terms that are closely related to software frameworks and design patterns are software composition and architecture.

Software architecture could be defined as "the structure of the components of a program/system, their interrelationships and principles and guidelines governing their design and evolution over time". [10] According to [10], recent emergence of interest in software architecture has been prompted by two distinct trends: a) development of methods, tools, languages and patterns for structuring complex software systems, b) exploiting specific domains to provide reusable frameworks for product families, based on idea of instantiating the shared design of common collection of elements for related systems. Principled use of software architecture can have positive impacts on software development in: understanding large systems, reuse, evolution, analysis and management, interoperability, market and integration with other non-software products. Contemporary approaches in architectural design in practice include: standardized components, product families, platforms, domain-specific architecture.

"Software architecture is a description of the way in which a specific system is composed from its components. Software composition is the construction of software applications from components that implement abstractions pertaining to a particular problem domain." [11]

Real benefit of "composable" software systems is in their flexibility [11]. "A flexible application can be achieved if architecture allows its components to be removed, replaced and reconfigured without perturbing other parts of the application. We see this phenomenon at work in object-oriented development: flexible classes of applications can be defined using frameworks, which are specified as hierarchies of reusable, abstract classes.", by providing libraries of reusable object classes. [11] Component-oriented development could be categorized at three levels:

- Framework level – "Framework is a set of reusable software components that forms the basis for an application." [1] "A generic software architecture is a description of a class of software architectures in terms of component interfaces, composition mechanisms and composition rules. A framework is a generic software architecture together with a set of generic software components that may be used to realize specific software architectures." [11] "Framework a set of classes that embodies an abstract design for solutions to a family of related problems. A framework is a partial design and implementation for an application in a given problem domain."

[22] Frameworks are concrete realization of groups of patterns that enable reuse of code[6].

- Composition level – specific applications are compositions of generic components from framework and new components obtained by specializing the generic ones [11].
- Instance level – A composition is instantiated to a running system, that may or may not be evolved (changed, modified) at this level. [11]

Design patterns provide methodological support for creating object-oriented frameworks. [11] [12] Design patterns are mechanisms for expressing object-oriented design experience. They provide a common vocabulary for design, constitute a base of experience for building reusable software, reduce system complexity, and they are building blocks from which more complex designs can be built. Design patterns identify, name, and abstract common themes in object-oriented design, capture the intent behind a design by identifying objects, their collaborations, and the distribution of responsibilities. [5] "A design pattern consists of three essential parts:

- An abstract description of a class or object collaboration and its structure.
- The issue in system design addressed by the abstract structure. This determines the circumstances in which the design pattern is applicable.
- The consequences of applying the abstract structure to a system's architecture. These determine if the pattern should be applied in view of other design constraints." [17]

3. Related Research

Different languages, tools and development environments, that support some degree of component-oriented development, define various kinds of components and notions in composition – no common model exists. Therefore, it is hard to describe common frameworks and architectures in a uniform way, to compare approaches or to analyze interoperability of data between software products created with different frameworks [11]. According to [10], software architecture-related research address issues such as: architecture description languages, formal models and mathematical foundations for modularization and system composition, architectural analysis techniques, architecture development methods, architecture recovery and reengineering etc. Special concern on architecture research is related to architectural decisions [13], styles [14] and evaluation [15].

The role, theoretical foundations and variations in applying software frameworks and design patterns has been explored in several theoretical and empirical studies [16, 17, 18, 19]. Importance of design patterns and frameworks for software development has been analyzed in [20]. Software reuse is particularly important within industrial approach to software development within "software factories" [21]. Industrializing software development leads to organizational changes that could be compared to other industries, such as using components from other code suppliers (ready-build or built-to-order components) and software supply chains. The key concepts of software development is abstraction (leads to model-driven development, meta-

models, transformations, patterns), granularity (component-based development), specificity (which introduces software product lines).

Different approaches, aspects and techniques of code reuse were explained within [1]. Approaches supporting software reuse include: application frameworks, application product lines, Aspect-oriented software development, component-based development, configurable vertical applications, CoTS (Commercial off the shelf) integration, design patterns, legacy system wrapping, program generators, program libraries, service-oriented systems. Reuse could be gained at different levels: code level components (modules, procedures, and libraries), entire applications, analysis level products, design level products. Code level reusable components could be built as: class libraries, function libraries, design patterns and framework classes. Special emphasize to programming language and development environments support to design patterns and frameworks is given within [22]. Meta patterns are used as the means for capturing the essentials in components and architectures that could lead to design patterns and frameworks. After being built upon some high-level abstract model, a software system could be evaluated and monitored during evolution, to examine if it tends to diverge from the intended and documented design model [24].

4. Existing Solutions Analysis

Existing solutions in software design patterns and frameworks could be generally categorized as object-oriented. Their applicability could be general or technology/semantics dependent.

4.1. Categorizing design patterns

As design patterns represent abstract solutions to a group of “software development situations”, they could be applied within different technologies and programming languages. Design patterns vary in their granularity and level of abstraction. There are many design patterns and they are categorized in aim to make it easy to refer to families or related patterns, to learn them or to find new patterns [17].

		Characterization		
		Creational	**Structural**	**Behavioral**
Jurisdiction	**Class**	Factory Method	Adapter (class) Bridge (class)	Template Method
	Object	Abstract Factory Prototype Solitaire	Adapter (object) Bridge (object) Flyweight Glue Proxy	Chain of Responsibility Command Iterator (object) Mediator Memento Observer State Strategy
	Compound	Builder	Composite Wrapper	Interpreter Iterator (compound) Walker

Figure 1. Categorization of common object-oriented design patterns [17]

4.2. Software frameworks solutions

Frameworks are developed as application of design patterns within particular technology as assembly of components. Therefore, frameworks are technology-oriented. Moreover, frameworks are developed as "generalized complete solutions" to a class of software applications and they could be instantiated with particular running software. The categorization of frameworks could also be semantic-based (giving general solution to software from certain applied domain).

"The first widely used framework, developed in the late 70's was the Smalltalk-80 user interface framework called Model/View/Controller (MVC). It devides user interface into three kinds of components: models, views and controllers. A model is an application object and it is supposed to be independent of the user interface. A view manages a region of the display and keeps it consistent with the state of the model. A controller converts user events (mouse movements and key presses) into operations on its model and view." [16]

A Web framework is a collection of software components that aids developers in the creation of Web-based user interfaces. [3] The framework manages the content displayed on the Web interface as well as which pages to display and what actions are available to the user of the page. A Web service framework is used to standardize the ways in which communication takes place between Web applications. [3] [4] Specific web services frameworks include SOAP (Simple object access protocol), RPC (Remote procedure call), REST (Representational state transfer). Common web applications architectural design patterns include multilayer, presentation-abstraction-control and model-view-controller [4].

An object-oriented communication framework is an integrated collection of components that cooperate to define a reusable architecture for communication software systems. [8] [10] Developing extensible communication software is a complex task. Increases in network and host performance require the design and implementation of communication software remains a challenging problem. Moreover, the growing heterogeneity of hardware and software architectures and diversity of operating system platforms make difficult reusing existing

algorithms, designs, interfaces, or implementations. Two techniques for overcoming communication software complexity and mentioned problems are design patterns and object-oriented frameworks. [8]

Multimedia streaming software is another application domain where software frameworks are commonly used. Multimedia frameworks are available on different operating systems e.g. Windows supports Direct Show, Linux provides GStreamer and Symbian enables Multi-Media framework (MMF). [6] Software design patterns are supported for information visualization, as well [25]. Some of design patterns include: proxy tuple, expression, scheduler, operator, renderer, production rule, camera design pattern and others.

Mobile devices are supported with specially designed web applications and services [3], as well as specific mobile applications. Mobile games are one of the most popular types of mobile applications and common design patterns are used for this particular purpose, such as MVC, state (state machine), singleton, observer. [26]

5. Conclusion

The importance of learning and applying industrial approach, methods and technologies to software development is supported by empirical results from many large companies. After introducing software design patterns and frameworks in their software production, they recorded 50-75% of improvements in productivity, software quality, customer satisfaction etc.

The goal of this paper was to systematize theoretical background, related research and existing solutions in software design patterns and frameworks. This paper contribution is in support more organized teaching and practice, as well as to encourage further research in this field. Related research is moving towards improving formal representation of software architecture, evaluation and software architecture design methods enhancements. Current practical solutions have evolved to large number of design patterns which are categorized as creational, structural and behavioral. Software frameworks are technology-oriented and support diversity of technologies such as web applications, mobile applications, communication software, multimedia etc. Of course, frameworks could be developed for certain application domains, such as healthcare, accountancy, higher education software etc.

Encouraging students to learn design patterns and frameworks brings them closer to software industry and raises their professional capabilities, so they could more easily adapt to professional practice. It is of a great importance for students to know basic principles and issues that lead to the need for design patterns, as well to encourage them to recognize new solutions and empower them to create their own solutions that could improve their analytical and critical abilities and creativity.

References

[1] Jalender B., Govardhan A., Premchand P., Designing Code Level Reusable Software Components, International Journal of Software Engineering & Applications (IJSEA), Vol.3, No.1, (2012).

[2] Vosloo, I., Kourie, D. G. Server-centric Web frameworks: An overview. Computing Surveys (CSUR), 40(2). (2008).

[3] Mohamed, K., Wijesekera, D. (2012). A Lightweight Framework for Web Services Implementations on Mobile Devices., IEEE First Internation Confernce on Mobile Services (MS), 64 – 71 (2012).

[4] Patrick C: Literature Review - Service Frameworks and Architectural Design Patterns in Web Development, University of Cape Town, (2014)

[5] Gamma E., Helm R., Johnson R., Vlissides J., Design Patterns: Abstraction and Reuse of Object-Oriented Design, ECOOP '93 Conference Proceedings, Springer-Verlag Lecture Notes in Computer Science (1993)

[6] Dajsuren Y, van den Brand M: Architectural and Design Patterns in Multimedia Streaming Software. Grace Technical Reports, pp. 31, (2009)

[7] Schmidtla D.C: Using Design Patterns to Develop High-Performance Object-Oriented Communication Software Frameworks. Proceedings of the 8th Annual Software Technology Conference,Salt Lake City,Utah, April 21–26, (1996)

[8] Schmidt D. C.: Experience Using Design Patterns to Develop Reusable Object-Oriented Communication Software. Communications of the ACM (Special Issue on Object-Oriented Experiences), vol. 38. (1995)

[9] Johnson R. Foote B: Designing Reusable Classes. Journal of Object-Oriented Programming, vol. 1, pp. 22–35, (1988)

[10] Garlan D, Perry D: Introduction to the special issue on software architecture. IEEE Transactions on Software Eng., (1995)

[11] Nierstrasz O, Meijler T.D: Research directions in software composition. ACM Computing Surveys, Vol 27, No 2, 262-264, (1995)

[12] Gamma E, Helm R, Johnson R, Vlissides J: Design patterns, Addison Wesley, Reading, MA, (1995)

[13] Jansen A, Bosch J: Software architecture as a set of architectural design decisions, Proceedings of the 5th Working IEEE/IFIP Conference on Software Architecture (WICSA'05),109 – 120, (2005)

[14] Monroe R.T, Kompanek A, Melton R, Garlan D.B: Architectural Styles, Design Patterns and Objects, Research Showcase, Carnegie Mellon University, (1996)

[15] Clements P, Kazman R, Klein M: Evaluating a software architecture. InformIT, (2001)

[16] Johnson R. E: Components, Frameworks, Patterns (1997).

[17] Gamma E, Helm R, Johnson R, Vlissides J: Design patterns, abstraction and reuse of object-oriented design, Proceedings of ECOOP ’93, Springer-Verlag Lecture Notes in Computer Science (1993).

[18] Johnson R.E: Frameworks=Components + Patterns. Communications of the ACM, Vol 40, No 10, (1997).

[19] Edwin N.M: Software frameworks, architectural and design patterns. Journal of software engineering and applications, Vol 7, 670-678, (2014)

[20] Ragnarsson O.A: Importance of design patterns and frameworks for software development. Reykjavik University, Reykjavik. (2014)

[21] Greenfield J, Short K: Software factories, assembling applications with patterns, models, frameworks and tools. Proceedings of OOPSLA ‘ 03, 16-27. (2003)

[22] Bosch J: Design Patterns and Frameworks: On the Issue of Language Support. Proceedings of the ECOOP '97 Workshops on Object-Oriented Technology, 133-136, (1997)

[23] Pree W: Meta patterns – a means for capturing the essentials of reusable object-oriented design. Proceedings of the European conference on Object-oriented programming ECOOP 1994, 150-162 (1994)

[24] Sefika M, Sane A, Campbell R.H: Monitoring compliance of software system with its high-level design models. Proceedings of the 18th International Conference on Software Engineering, IEEE CS Press, 387–396, (1996)
[25] Heer J, Agrawala M: Software design patterns for information visualization. IEEE Transactions on visualization and computer graphics, Vol 12, No 5 (2006).
[26] Ilja A: Use of design patterns for mobile game development. Bachelor thesis. Umea Universitet (2012).

Hadoop as a Platform for Big Data Analytics in Healthcare and Medicine

Blagoj Ristevski[1], Monika Stevanovska[2] and Borche Kostovski[3]

Faculty of Information and Communication Technologies - Bitola
"St. Kliment Ohridski" University – Bitiola, R. Macedonia
Partizanska bb 7000 Bitola, Republic of Macedonia
[1] blagoj.ristevski@fikt.edu.mk
[2] monikastevanovska@gmail.com, [3] borchek@hotmail.com

Abstract. Big data analytics has the potential to alter the way that life sciences and healthcare and medical organizations use sophisticated technologies to make appropriate decisions. Big data analytics allows companies and organizations to examine and explore large and various data and thus to identify relationships, trends, and patterns to reveal insights, comprehensions and knowledge hidden in the data. In order to discover the knowledge from big data several challenges have to be addressed. As big data analytics becomes more ordinary process in healthcare organizations, concerns regarding patients' privacy, data security, establishing standards and good governance policy should be made. Moreover, improving the big data analytics tools is required to be able to resolve computational complexity issues in a cost-effective manner. In this paper we suggest using of Hadoop as a suitable platform for big data analytics in healthcare and medical organizations.

Keywords: Big Data Analytics, Healthcare Information Systems, Bioinformatics, Medical Informatics, Database Systems.

1. Introduction

Big data is a term that describes the large volume of structured, unstructured and semi-structured data that are produced on a daily basis. Big data analytics denotes the process of examining varied and huge amount data sets i.e. big data in order to reveal hidden patterns, unknown correlations, trends, customer habits and many other useful information, which can help organizations to make more up-to-date decisions.

Nowadays, healthcare and medical organizations and companies generate and collect large amount of heterogeneous data such as data from medical sensors, magnet resonance imaging, computer tomography, x-ray imaging, patients' personal data, prescriptions, as well as different –omics data: genomics, proteomics, metabolomics and transcriptomics data. These data are stored in diverse data formats and their analyzing is very challenging task.

Hadoop is an open-source software framework for storing data and running applications on clusters of hardware. It provides massive storage for any data types, enormous processing power and the ability to handle almost limitless concurrent tasks virtually [7].

The rest of the paper is structured as follows. Section 2 describes the term big data and its characteristics. The next section depicts usage of big data in the healthcare and medicine and the analytics framework. Section 4 describes the most popular framework for big data analytics – Hadoop, its algorithm and benefits of its application. The relationship between Hadoop and healthcare, as well as the need of Hadoop in the healthcare data solutions are shown in the Section 5. Section 6 depicts the electronic health records, while last section provides concluding remarks and some direction for further works.

2. Big Data

Big data describes the large volume of various data: structured, semi-structured as well as unstructured data that are generated every days. It is not important only the amount of data, but what the companies and organizations do with these collected and stored data. Big data can be analyzed for insights that lead to better decisions and strategic business changes.

Every part of business and society are always changing due to that fact that there are so huge amount of data and hence there is a need to develop suitable software packages and tools for data analyses. But, what are the properties of big data? Big data is commonly characterized using a number of V`s [8]. Fig. 1 depicts the most common properties of big data so-called 4V's of big data:

- **Volume** - the quantity of generated and stored data. The size of the data determines the value and potential insight.
- **Variety** - the type and nature of the data that are heterogeneous.
- **Velocity** - the speed at which data is generated and processed.
- **Veracity** - the quality of the collected data can vary significantly.

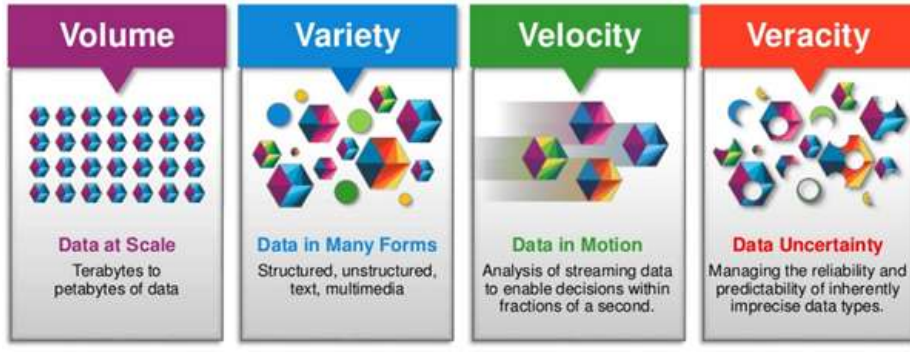


Fig. 1 The 4 V's big data properties: volume, variety, velocity, veracity [9].

3. Big Data in Healthcare and Medicine

Data analytics in healthcare and medicine is a symbiosis of clinical innovation and modern information and communication technology. As the healthcare industry continuously generate large amounts of data in different data formats, it is almost impossible to handle these data stored in various data formats. Certainly, the current trends favor digitization of these large amounts of data to be suitable for further data analysis.

Nowadays, big data analytics offers a new way to healthcare and medical organizations and companies to develop actionable insights, to organize their future directions, to boost up their decision making and to gain suitable results worthwhile. Big data analytics emerges as an advanced leading analytic technique that analyze and manage the large amount of digitalized data. This approach is getting superior and better as numerous companies have competed in the contest to provide the best healthcare analytics services.

Data analytics in healthcare and medicine can be used to raise the standards in the following fields: public healthcare, electronic medical records (EMRs), patient profile analytics, genomic analytics, fraud analysis and safety monitoring **Error! Reference source not found.**.

The conceptual framework for a big data analytics project is similar to those projects used in standard business intelligence or analytics. The main difference is in the manner of carrying out of the processing. In a standard analytics project, the analysis can be performed with a business intelligence tool installed on a stand-alone system such as a desktop or laptop.

While the concepts of distributed processing are not new and have been existed for decades, their use in analyzing huge amount of data sets is relatively new as companies start to store their data repositories to gain insights to make more useful and informed decisions. Additionally, the availability of open-source platforms such as Hadoop/MapReduce on the cloud has further stimulated the application of big data analytics in various areas [11].

4. Hadoop

The most significant platform for big data analytics is the open-source distributed data processing platform Hadoop (Apache platform), initially developed for routine functions such as aggregating web search indexes [12]. It belongs to the class of NoSQL technologies that have evolved to aggregate data in unique ways, while others include CouchDB and MongoDB. It is generally accepted that there are two important modules in Hadoop: Hadoop Distributed File System (HDFS) and MapReduce [13].

MapReduce is a programming framework developed by Google that supports the underlying Hadoop platform to process the big data sets residing on distributed servers (nodes) in order to produce the aggregated results. MapReduce algorithm uses the following three main steps: map function, shuffle function and reduce function [14].

The main benefits of using Hadoop are its highly cost-effective properties, great data reliability [15]. Hadoop provides extremely scalable solutions, that are simple, fast and flexible, and moreover it offers comprehensive authentication and data security.

The conceptual framework of big data analysis project in health care is similar to that of a traditional health analytics project. A system architecture is shown on Fig. 2. The main difference between both lies in how processing is executed. In a regular health analysis project, the analysis can be performed on a standalone system, such as a desktop or laptop.

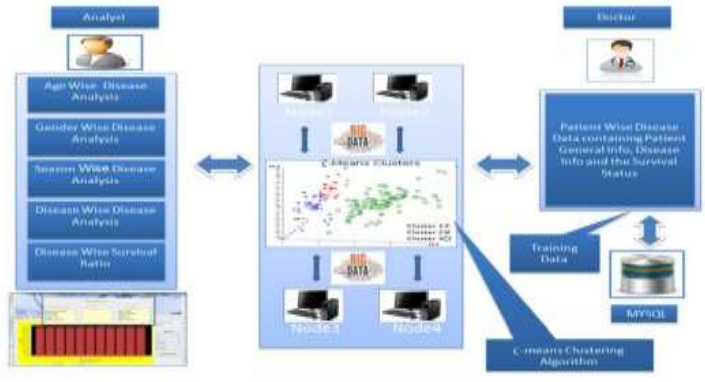


Fig 2. System Architecture [16].

5. Hadoop and Healthcare: A timely encounter

Hadoop can play a main role as the great aggregator of the large and heterogeneous collections of data collected from various sources.

As evidence of disaggregation the health records of an individual is fundamentally fragmented by feature of the many groups of EHRs that a person might encounter over a lifetime. By combining these partials data collections together, a more complete record can be obtained. Integration of these diverse data enables organizations and companies to solve kinds of issues, such as those in the category of healthcare management. Some of these issues cannot be solved by using standard methods of integrating distinct data systems.

Another key implications for hospitals are the financial implications especially for the readmission patients. This means that the healthcare providers have to take into account prevention for those patients with highest risk of readmission. This problem is solved using a predictive model constructed from machine learning methods whose training data is drawn from various EHR sources. Another captivating use case is predicting the onset of different kinds of diseases by using a predicting model with data ideally flowing from multiple EHR sources. However detection algorithm is time-sensitive and have to be deployed into a near real-time monitoring environment [17].

6. Electronic Health Records (EHRs)

The electronic health records (EHRs) itself could be considered as "big data" and hence extend to the handling and application of data stored in EHRs.

Over the years, EHRs data have been used with the intent to improve care, to increase patient engagement, perform quality improvement, build shared models and standardization across institutions, create new knowledge, carry out research in a "real-world" settings instead of in controlled conditions. These data enable public health investigation and simplify personalized care and decision making in healthcare and medicine [18] [19]. The final goal is to create a continually learning healthcare infrastructure with real-time knowledge production and create a predictive, preventive and personalized healthcare system [20].

Employing the EHR system to answer healthcare questions varies from the traditional research approaches of collecting data. Although EHRs have been in existence for many years, many challenging tasks exist. These challenges include limitations of processing ability, interoperability and lack of standardization, accuracy and completeness of records, costs, concerns in ability to extract the needed information, as well as security and privacy of the patients' data [21] [22].

7. Concluding remarks and further works

The aim of modern healthcare systems is to provide optimal healthcare by using of healthcare and medical information technology in order to improve healthcare quality, so that outcomes are consistent with current professional knowledge; to reduce healthcare costs and to provide support for healthcare system [23].

The so-called Industrial Internet, which refers to the rapidly increasing number of smart, interconnected devices and sensors generates volumes of data. Nowadays, a variety of devices monitor every sort of patient behavior, such as from glucose, blood pressure and electrocardiograms monitors. The data from these various monitors can be analyzed in real time and send alerts to care providers so they know instantly about changes in a patient's condition. Processing real-time data with machine learning algorithms can provide clinicians with insights to help them make decisions and allow for effective interventions. Wearable sensors and devices present the opportunity for interactions with patients, making healthcare more convenient.

The cost of fraud, waste, and abuse in the healthcare industry is a key contributor to increase healthcare costs. To identifying fraud and waste, analyze large unstructured datasets of historical claims and to use machine learning algorithms to detect anomalies are needed.

Regarding above mentioned concerns and issues in healthcare and medicine, Hadoop offers an appropriate platform for development of big data analytics solutions.

References

[1] Cortada, J. W., D. Gordon, and B. Lenihan. "The value of analytics in healthcare: from insights to outcomes. IBM Global Business Services." *Life Sciences and Healthcare, Executive Report* (2012).
[2] Institute of Medicine, "Shaping the Future for Health", Institute of Medicine, 1999, pp. 1-8
[3] Coiera, E., Walther, M., Nguyen, K., and Lovell, N.H., "Architecture for Knowledge-Based and Federated Search of Online Clinical Evidence", Journal of Medical Internet Research, 7(2005, pp. e52.
[4] K. Hayrinen, et al., "Definition, structure, content, use and impacts of electronic health records: A review of the research literature," International Journal of Medical Informatics, vol. 77, pp. 291-304, May 2008.
[5] Taylor, R. C., An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics. BMC Bioinforma. 11:6, 2010.
[6] Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., Morton, S. C., and Shekelle, P. G., Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann. Intern. Med. 144(10):742–752, 2006.
[7] https://opensource.com/life/14/8/intro-apache-hadoop-big-data
[8] https://www.mindtory.com/introduction-big-data/
[9] https://pro2col.com/challenge-big-data-more-than-just-big-files/
[10] Lee, Choong Ho, and Hyung-Jin Yoon. "Medical big data: promise and challenges." *Kidney research and clinical practice* 36.1 (2017): 3.
[11] http://www.smartdatacollective.com/5-ways-hadoop-can-help-healthcare-organizations-and-you/
[12] http://ercoppa.github.io/HadoopInternals/HadoopArchitectureOverview.html
[13] https://www.tutorialspoint.com/hadoop/hadoop_mapreduce.htm
[14] https://www.journaldev.com/8848/mapreduce-algorithm-example
[15] https://www.technavio.com/blog/top-5-benefits-using-hadoop
[16] https://www.irjet.net/archives/V2/i7/IRJET-V2I788.pdf
[17] https://hadoop.cioreview.com/cxoinsight/hadoop-and-health-care-a-timely-encounter-nid-11843-cid-154.html
[18] The promise of comparative effectiveness research. *Sullivan P, Goldmann D JAMA. 2011 Jan 26; 305(4):400-1.*
[19] The economic value of health care data. *Harper EM Nurs Adm Q. 2013 Apr-Jun; 37(2):105*
[20] A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. *Hood L, Flores M N Biotechnol. 2012 Sep 15; 29(6):613-24.*
[21] Big data, bigger outcomes: Healthcare is embracing the big data movement, hoping to revolutionize HIM by distilling vast collection of data for specific analysis. *Fernandes L, O'Connor M, Weaver V J AHIMA. 2012 Oct; 83(10):38-43; quiz 44.*
[22] Defining and measuring completeness of electronic health records for secondary use. *Weiskopf NG, Hripcsak G, Swaminathan S, Weng C,J Biomed Inform. 2013 Oct; 46(5):830-6.*
[23] https://mapr.com/blog/5-big-data-trends-healthcare-2017/

Short Review of Cloud Computing

Edith Boral [1], BSc Nebojša Tatomirov [2], prof. dr Ivana Berković [2], MSc Nikola Petrović [2]

[1] Archbishop Molloy HS, Briarwood NY, USA
{editboral@hotmail.com}
[2] Technical faculty "Mihajlo Pupin", Đure Đaković bb, Zrenjanin, Serbia
{nebojsa.tatomirov@tfzr.rs, berkovic@tfzr.uns.ac.rs, petrovic.n26@gmail.com}

Abstract. Cloud Computing represents code technology where IT experts striving towards to increase capacity and add new capabilities to systems already in use without any extra costs in new infrastructure and hiring new people. Cloud computing is considered as an architecture, platform, operating system and service. Cloud computing is considered as cloud structure in "cloud" where business and private users can use application from anywhere in the world. The capacity offered to the users is large, but not infinite.

Keywords: Cloud computing, Amazon, Google, Microsoft.

1. Introduction

Cloud computing has greatly contributed to the IT industry from its inception until today. It is based on five key features that show the relationship and difference between cloud computing systems and traditional approach to the computers. Virtualization has a big role, it also contributes to the quality and it makes it easier to use cloud computing.

2. Definition of Cloud Computing

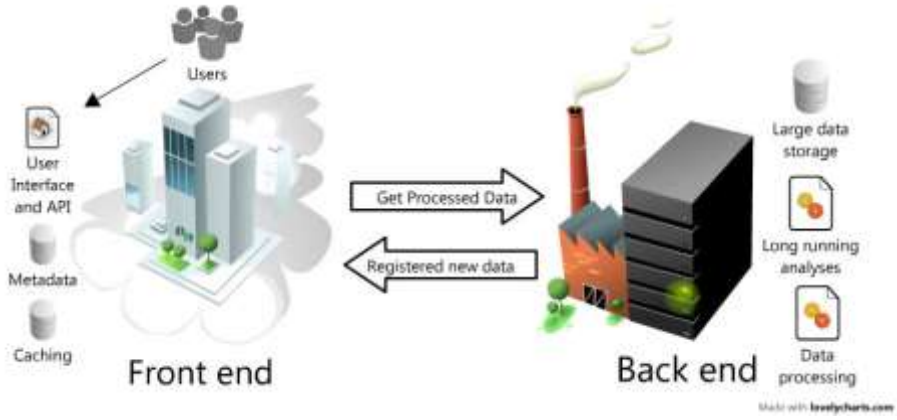
In last few years, the popularity of cloud computing is getting bigger and bigger, while the idea of their development is from the sixties of the last century. One of the main reasons for cloud computing popularity is reducing the costs of hardware, computer and network resources needed to maintain the cloud computing infrastructure without any difficulty. The number of cloud computing services has increased due to small investments in equipment, which has led to an increase the number of available applications through the mentioned infrastructure.

One of the main definition of cloud computing says that it represents computer model where services are placed on the internet, where users can access them with proper conditions. A more complex definition says that cloud computing services represent computers and programs set where to it services are accessed via the internet.

Picture 1. Cloud computing

2.1. Cloud Computing Architecture

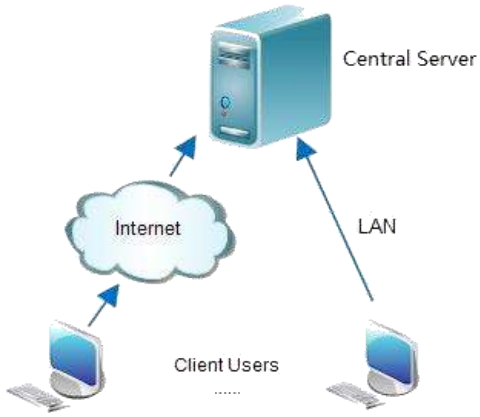
The division is divided into two parts: front end and back end. They are interconnected through the network, most often via the internet. Front-end represents the user or client side, while the back end represents the cloud section of the system.



Picture 2. Front-end and back-end connection

Front-end includes user computer and applications necessary for cloud computing access. The interface of cloud computing services is different. Services as web based e-mail programs use already existing web browsers like Google Chrome, Firefox,

Safari and many others. Other systems have unique applications that provide network access to the user.



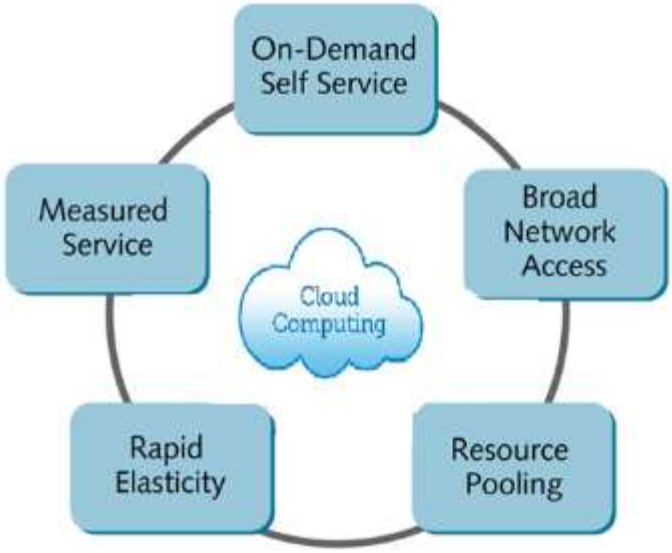
Picture 3. Cloud computing central server

Most of the time, servers don't work with full capacity, it means there is unused processor power. There is an option to cheat physical server to act as multiple servers, where each of them works and execute its own operating system. That technique is called server virtualization. Increasing results of an individual server, server virtualization reduces the need for more physical computers. In theory, cloud computing system can include any computer program, from date to video game.

2.2. Cloud computing characteristics

Difference between cloud computing systems and traditional approach to computing can be described with five basic and key characteristics.

- On-demand self-service – The user independently selects and run computer resources. The user can choose the time of serving and network place to storage his date without any interaction with employers in cloud computing company.
- Broad network access – Cloud computing possibility can be accessed over the network.
- Resource pooling – The provider's computer resources are merged to satisfy all users, using the model of multiple leased units (Multi-Tenant model).
- Rapid elasticity – Cloud computing possibility can be improved and elastically started. Usually, this process is automated for better proportional increasing or decreasing possibility when they are no longer needed.
- Measured service – All systems that use cloud computing execute automatic check and optimize resource usage.



Picture 4. Cloud computing characteristics

3. Comparing Amazon, Google and Microsoft Cloud computing services and their characteristics

Picture 5. Cloud computing liders

3.1. Storage

Storage on cloud works completely different from normal storage on a hard drive or solid-state drive. Systems most found solution for multiple problems, also it needs to provide security in data transfer, that is, not to allow any information to be lost.

The best solution has Amazon, but neither Google nor Azure is not too much behind it. Amazon S3 allows you to analyze data directly on the cloud. Amazon is the most advanced storage platform with the largest ecosystem of solutions.

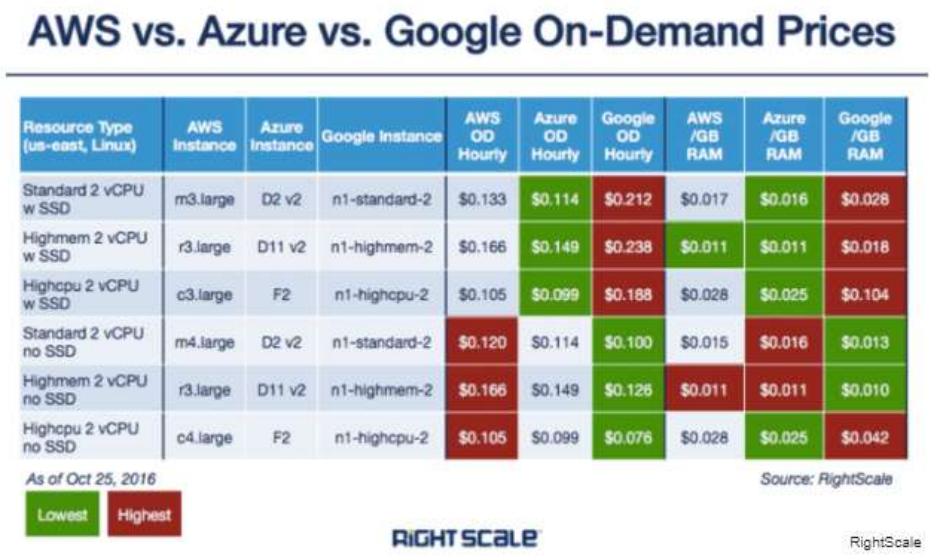
All three systems are very reliable. In line with the offer, this service also has its price. In addition to having the best solution, Amazon is also the most expensive solution. All three cloud computing systems have backup solutions differently, depending on the purchased package.

3.2. Databases

When it comes to databases, the lack of options that Google offers to the end user gives a huge minus to this system, while Azure and Amazon offer many options that make it easy for users to use and work on the system. Amazon RDS (Relational Database Service) provides support to every database including MySQL, Oracle, PostrgeSQL, etc. This system manages with updating options, as well as helping to the user to solve common problems with the database.

3.3. Price and comparison of costs

There are many of factors that affect on the price of cloud computing service, the size of virtual machine, type of the virtual machine, etc.



Picture 6. Price comparison of Cloud computing services

It is very difficult to decide which manufacturer has the most expensive virtual machine because there are many different options for these three providers.

When a user wants a multi-year contract, Google usually has the lowest price for a virtual machine, while Azure usually satisfies the user's needs and surpass the price offered by Amazon.

4. Conclusion

Cloud computing is a technology whose number of users is constantly growing and therefore this technology constantly evolves. Lately, there are more "ordinary" users who begin to see the benefits of cloud computing, for economic reasons, saving money to buy external devices such as USB flash drives, CD, etc. Cloud computing develop tends to increase capacity with little or no cost to invest in infrastructure. Security problems are present the biggest obstacle in cloud computing, but despite the problem, cloud computing has a bright future due to the maximum use of computing resources.

5. References

1. Anthony T.Velte, Toby J.Velte, Ph.D.Robert Elsenpeter - „Cloud Computing: A Practical Approach", McGraw Hill, 2010
2. Dimitris N. Chorafas – „Cloud Computing Strategies", CRC Press, 2011
3. Barrie Sosinsky - „Cloud Computing Bible", Willey Publising Inc, 2011

ICT in Education in Montenegro and Serbia

Arben Lunjić[1], and Savina Čolić[2]

[1] Primary school "Boško Strugar", Ulcinj, Montenegro
[2] University of Novi Sad, Technical faculty "Mihajlo Pupin"
Zrenjanin, Serbia
savina.dj@gmail.com

Abstract. It is widely known that the educational systems of countries in South-Eastern Europe are rather traditional, not following the trends from the developed countries and needs of contemporary economic environment. This tardiness has been identified in numerous ways, and certain steps have been made towards improving this situation. Strategies for development of information society, including the education development in this region, have been introduced in the beginning of the second decade of the 21st century, with the aim to modernize the process and help young generations develop skills needed in today's business world. This paper will give an overview of the current and desired state of educational systems of Montenegro and Serbia, respectively, and provide a critical thought and guidelines for their further development.

Keywords: information and communication technology, information society development, education development, Montenegro, Serbia.

1. Introduction

ICT enable control and facilitate various aspects of our lives. Full mobility, houses responsive to the needs of their inhabitants, intelligent cars, health and environment monitoring and problem prevention, are just a fraction of opportunities offered by the use of ICT. These capabilities will soon become generally available, introducing a completely different use of these technologies. The ICT and related network and service infrastructure do and will have a high impact on various aspects of life - ageing and inclusion, business productivity resulting in economic growth, and as such, they should generally be recognized as a means for achieving better situation in society.

On the other hand, current situation with economic downturn and focus on short term reactive measures at policy-making level, poses a challenge where ICT innovations can play a significant role. Combined with the previously mentioned potential, this setup requires a systematic approach which provides young generations with the knowledge and technical capabilities to develop and frame the new Internet infrastructure, services, access devices and systems. Although the young generations grow surrounded by

modern technologies, they should be exposed to it through systematic and guided approach, offered at schools.

Therefore, greater attention should be paid to the ways ICT is presented to and implemented in school system, creating the basis for the future. In this light, numerous strategies have been introduced and utilized at various levels for over a decade now. This paper shall present the overview of the ICT in education in Montenegro in parallel with the same aspect in the republic of Serbia.

2. Information Society Development

The trend of developing open, competitive, advanced and secure information societies throughout the world is present in the South-Eastern Europe as well, with the main units of the educational institutions equipped with the information technology. Along with the other areas of working life, education supports the development of 'a well-functioning internal market for electronic communications, electronic commerce, and audio-visual services', as stated in the European Commission report on Montenegro's progress in 2015 [1].

Certain positive effects of the development are already present, but the more thorough analysis, done for the purposes of developing the strategies for the rise of Information Society, show that there is still a great deficit of the ICT skills in this region, significantly slowing down the multiplicative effect of the modern technologies on productivity growth. Some of the effects are development of commerce, enhanced use of the capital, and strengthening the national competitiveness. Becoming an integral part of the national infrastructure, in parallel with the energy and transport, the ICT have become a significant factor of production in most segments of the modern economy.

2.1. Snapshot of the Current State

The situation as seen nowadays is marked by everchanging environment, including labor market largely fluid by nature, requiring creativity, team work, problem solving, critical thinking and other skills that the traditional educational systems cannot develop.

Moreover, the demand for advanced ICT skills is also rising due to the increased reliability of the modern economy on the Internet and communication technology. Combined, these two facts mean that the education should be structured in such a way that it stimulates the development of the aforementioned skills. In this way, the education and training should be modernized from the curriculum, over the evaluation of the acquired knowledge, to the professional development of teachers.

The state as is, recognizes the e-education as not sufficiently represented in the educational system of the Member States.

Currently, the Strategy for the Information Society Development 2020, outlining the strategic development tools in this development field, with the goal to reach the EU

standards set out in the Digital Agenda 2020 and the Digital Single Market Strategy, is positive and largely implemented in Montenegrin educational system. The Chapter of this Strategy, important for this paper, is the Chapter on E-Education.

One of the problems faced by the countries trying to tackle the distribution of the Internet access is the new digital gap – introducing the uneven distribution of the skills.

2.2. Strategy for the Information Society Development

Currently, the Strategy for the Information Society Development 2020, outlining the strategic development tools in this development field, with the goal to reach the EU standards set out in the Digital Agenda 2020 and the Digital Single Market Strategy, is positive and largely implemented in Montenegrin educational system. The Chapter of this Strategy, important for this paper, is the Chapter on E-Education.

The identified benefits of e-education are as follows: acquiring new skills, new teaching methods, individualization of learning and teaching methods and styles adjustable to the situation and students' and teachers' needs, and lowering of the training price and time efficiency.

However, one of the problems faced by the countries trying to tackle the distribution of the Internet access is the new digital gap – introducing the uneven distribution of the skills due to the unavailability of technologies to certain areas of the world. This problem is also tackled by the Strategy. This largely distinguishes the students who have access to the innovative education from the ones who are marginalized in this way.

In last instance, the number of jobs requiring ICT skills is growing, with not only the commercial market, but also the public sector being gradually transformed into e-government. Other areas increasingly facing the change towards the electronic means are research and development, healthcare system, education and such. This is the reason why the priorities of the Strategy include the increase of formal and informal certification of the ICT graduates and increase of the share of the ICT in GDP, reflected in economic growth and job creation in other sectors of the economy.

2.3. E-Education in Montenegro

The current state is still not perceived as completely positive, with very low level of ICT provided in the educational institutions of all levels and thus low chances provided. Due to this, the focus is to increase this proportion of computers to students, the skills of the teaching staff as well as the level of the development of the digital materials for teaching and learning, like didactic software, simulators and similar. Unfortunately, little attention is paid to the importance of the digital competences in the information society, despite the fact that a large proportion of job positions require workforce with acquired digital skills. Finally, a worryingly small number of staff educated in this area contributes to the problem as well.

Montenegro has introduced the Strategy for the information society development, with the final goal here to eliminate the digital divide between the urban and rural areas, bridging the gap between the social and demographic characteristics influencing the economic benefits for each Montenegrin citizen. To achieve this, a lot of effort still needs to be put into the process, starting from enhancing the internet coverage, over equipping the educational institutions with ICT and finally, improving the knowledge and skills of the teaching staff.

Strategic indicators, as provided in the Strategy for the Information Society Development 2020 [2], shows the targeted trend as presented in the following Table:

Table 1. Strategic indicators of the development

Indicator	Current state	2018	2020
Student-computer ratio	1:16 (elementary schools) 1:14 (secondary school)	1:12	1:10
Internet coverage in regional school units	0%	50%	100%
Internet speed in main units	4 mbps fixed 8 mbps satellite	8 mbps fixed	20mbps fixed
Percentage of teachers trained (basic IT skills)	20%	25%	30%
Percentage of teachers trained in IT security	1.50%	10%	20%
Percentage of use of teachers' web portal	1%	10%	20%

2.4. E-Education in Serbia

The state as is in Serbia is somewhat more favorable. In order to get an insight into the current state in Serbian educational system, the positive legislature and acts shall be investigated. Just like in Montenegro, Serbia has also introduced the Strategy for Education Development in Serbia 2020. Introduced in 2012, this document cites knowledge and creativity as prerequisites for yielding long-term progress both in economy and life, even in the world of uncertainties. This document sees various forms of activities as the online form of education. They include e-conferences, subject blogs, e-testing and similar. Moreover, it suggests the use of various forms of distance learning as one of future trends.

Another document interesting for this topic is the National Youth Strategy, implemented in 2008. It cites the integration of Serbian educational system into European educational space as its imperative. Moreover, as presented by this strategy, it is very important to develop a mechanism to increase the e-literacy of the young population. Moreover, as stated by the strategy, the command should be handed from the teachers to the students. Just as in Montenegro, the education process is oriented to each individual student, with the final goal to adjust to the individual needs of each student to the maximal level, which is impossible to achieve in classical educational process. This learning not only provides greater autonomy, but also active learning, where the student creates, communicates and participates actively, up to the point where the boundary between the teacher and student is removed.

3. Conclusion

As it could have been seen, although slowly, our region did start its journey towards innovation of educational system, including the greater utilization of ICT. There are still numerous improvements to implement, starting from the infrastructure, over education of teachers and reforming the curricula, but the initial steps have been taken. The fact that the ICT are very useful in present and future education is clear and well-known, and will see the more extensive use in the future.

If approached to strategically, this strive can bring upon multiple benefits, ranging from development of skills needed on today's market, increasing the quality of education through tailoring it to the needs of every situation and participants in it, and finally, reducing the costs of the process. Since the educational system is proved to be lagging after the workforce market needs significantly, the pace of its development should be increased, but not at the cost of quality.

To sum up, the Strategies presented in this paper should serve as the guidelines, but the process should be enforced to greater lengths in order for our region to become more competitive on the wider marketplace.

4. References

1. Montenegro 2015 Report, European Commission: Commission Staff Working Document, Brussels, Belgium (2015)
2. Strategy for the Information Society Development 2020, Montenegro Ministry for Information Society and Telecommunications, Podgorica, Montenegro (2015)
3. Revising Europe's ICT Strategy, Report from the Information Society Technologies Advisory Group (ISTAG), final version (2009)
4. Strategy for Education Development in Serbia 2020, The Ministry of Education, Science and Technological Development of the Republic of Serbia, Belgrade, Serbia (2013)
5. Čolić, S.: The High School of Economics Students' Opinion on Using the Information Technologies and Internet in Schooling (Master's Thesis), Novi Sad, Serbia (2015)

Causality library for formal system

Aleksandar Berar, Višnja Ognjenović and Ivana Berković

University of Novi Sad, Technical faculty "Mihajlo Pupin"
23000 Zrenjanin, Serbia
aleksandar.berar@gmail.com

Abstract. Java programming language is used fairly often nowadays. However, due to the fact that it is a compiled programming language, Java carries certain problems. The main one is the fact that if the code needs to be changed, the whole project has be re-compiled in the majority of cases, with artifacts being reinstalled. The problem as described, is solved by introducing a library named Causality, which makes it possible to isolate parts of the system to be redefined. This paper will present the library, its setting, features, classes, and finally give a simple example of the implementation of Causality library in practice.

Keywords: Java, Causality library, formal system

1. Introduction

In this paper a java library called Causality will be presented. This name is chosen because the effect of the library resembles that term; the predetermined predicate chooses a subclass of a superclass which corresponds to the effect or consequence. For example, if there is a superclass called Shape, it is split into classes like Triangle, Square, etc. then predicates can be defined from which it would follow the appropriate class selection.

Java programming language is a compiled programming language [1]. So if there is a need to make a change to the code, often the whole project must be re-compiled and artifacts must be installed again. Because of this case and problem, this library was created. In this way, it is possible to isolate parts of the system that can be redefined by altering the textual information consisting of JSON textuality [2] and defining predicates and selecting class implementations in a different way [3], thereby avoiding re-compiling the entire project.

2. Mathematical setting

A formal system or logical calculus is any well-defined system of abstract thought based on the model of mathematics. A formal system need not be mathematical as such. Each formal system uses a set of primitive symbols (sometimes known as an alphabet) to finitely construct a formal language from a set of axioms through inferential rules of formation. The system thus consists of valid formulas built up through finite

combinations of the primitive symbols—combinations that are formed from the axioms in accordance with the stated rules [4]. Mathematical setting of Causality system is shown in Equation 1.

$$c(X) = \begin{Bmatrix} p_s \wedge p_1(x_1,x_2,\ldots x_n) \Rightarrow K_1, \\ p_s \wedge p_2(x_1,x_2,\ldots x_n) \Rightarrow K_2, \\ \vdots \\ p_s \wedge p_m(x_1,x_2,\ldots x_n) \Rightarrow K_m, \\ \text{otherwise} \Rightarrow K_o \end{Bmatrix} \quad (1)$$

- X is the alphabet of arbitrary variables.
- $x_1,x_2,\ldots x_n$ are elements of the alphabet, or symbols.
- $p_1,p_2,\ldots p_m$ are predicate functions that translate the set of arbitrary variables into the set of truthfulness.
- otherwise is the negation of disjunction of all predicate functions.
- K is the superclass.
- $K_o,K_1,K_2,\ldots K_m$ are the subclasses.
- $K_m \subseteq K$ is a relation that represents "Km is a subtype of K".
- p_s is the super-cause.

3. Classes of Causality system

In further subchapters of this chapter, classes of Causality system will be presented. Classes of this system are split into three modules: Predicate package, Matcher package and Causality package. All of these three submodules are part of the higher order module which represents entirety of Causality library which is packed as a Java archive file [5][6].

3.1. Predicate classes

The basis of the predicate class packet is the Predicate interface, which consists of two methods: test and flattened. The test method is an important method that returns a Boolean value for a predicate. Flattened method is a method that returns the stream of predicates since some predicate implementations contain multiple predicates.

Condition is a typed class that implements the Predicate interface. It is slightly more complicated than other classes that implement this interface. One field of this class is Matcher - a matching class that will be explained in the next chapter. It is important to note that the Matcher type corresponds to a type that defines a typed object of the Condition class. The objects of this class are practically used to compare the values of the key variables with a parameter over the matching Matcher class.

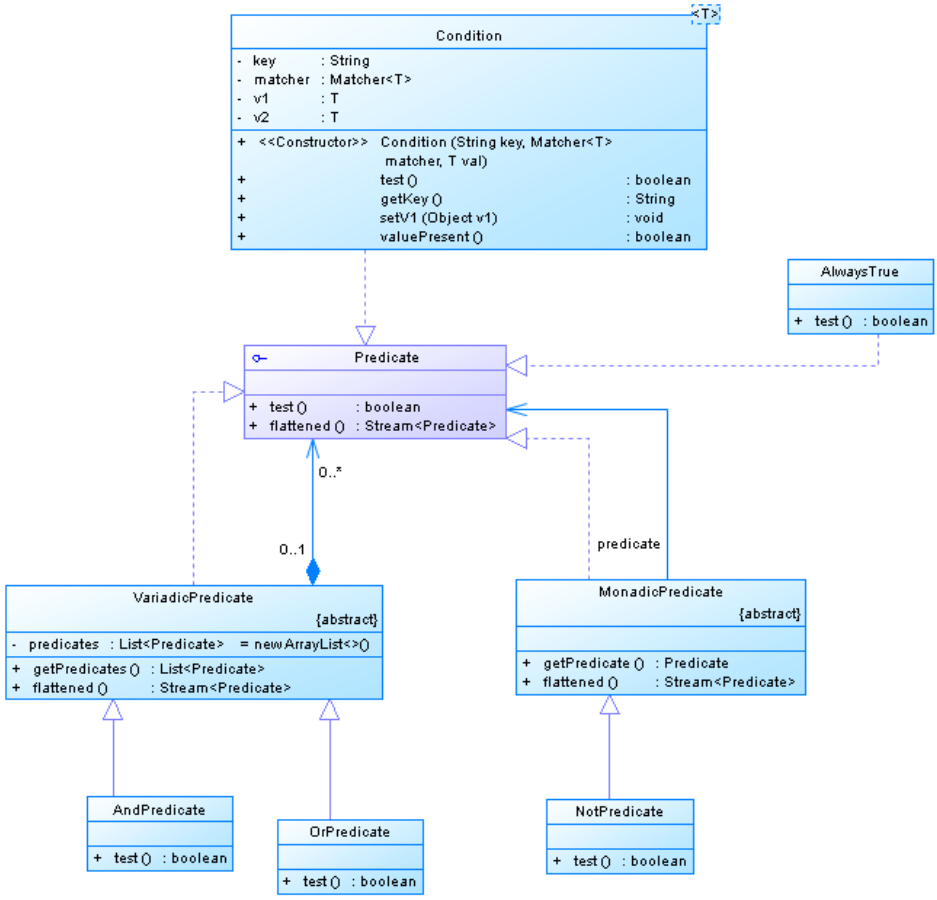


Fig. 1. Hierarchy of Predicate classes

```
{¶
··"type":·"cond",¶
··"key":·"b",¶
··"matcher":·{¶
····"type":·"string",¶
····"impl":·"equals"¶
··},¶
··"value":·"test"¶
}¶
```

Fig. 2. Example of Condition class defined as JSON text

3.2. Matcher classes

Basically, the Matcher interface is a typed interface and the idea is that based on two input values that are of this type, it returns a Boolean type as the output value.

Matcher interface implementation is an abstract typed AbstractMatcher class that contains a private field from an associative array matchers whose key is a String object that points to a Matcher of type T. This associative array will be filled by classes that inherit the AbstractMatcher class. These are the NumericalMatcher classes and StringMatcher.

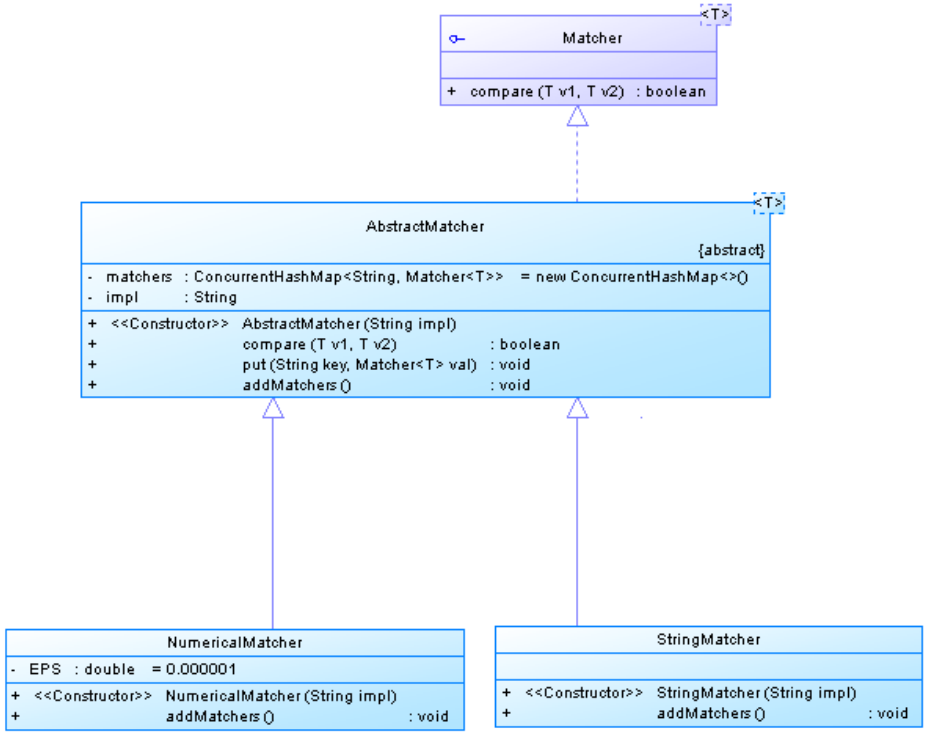


Fig. 3. Hierarchy of Matcher classes

```
public class NumericalMatcher extends AbstractMatcher<Double> {
    private static final double EPS = 0.000001;
    @JsonCreator
    public NumericalMatcher(@JsonProperty("impl") String impl) {
        super(impl);
    }
    @Override
    public void addMatchers() {
        put(">", (v1, v2) -> v1 > v2);
        put(">=", (v1, v2) -> v1 >= v2);
        put("<", (v1, v2) -> v1 < v2);
        put("<=", (v1, v2) -> v1 <= v2);
        put("==", (v1, v2) -> Math.abs(v1 - v2) < EPS);
    }
}

public class StringMatcher extends AbstractMatcher<String> {
    @JsonCreator
    public StringMatcher(@JsonProperty("impl") String impl) {
        super(impl);
    }
    @Override
    public void addMatchers() {
        put("equals", ((v1, v2) -> v1.equals(v2)));
        put("startsWith", ((v1, v2) -> v1.startsWith(v2)));
        put("endsWith", ((v1, v2) -> v1.endsWith(v2)));
    }
}
```

Fig. 4. Implementations of numerical and string matchers

3.3. Causality classes

Causation is a typified class that is the most important class in this library. Each Causation object consists of a superCause field that is of the Predicate type and this predicate corresponds to an over-condition that must always be fulfilled for the given problem. The next field is the causalities that is a list of Causality objects of type of T. The field otherwise is a field of type T and corresponds to the result or effect when no predicate from the predicate part within the causalities list is met.

It is important to note the valueMap field that is of type of associative array with the keys that are String objects and values that are Object objects themselves. Within this field, the values of the variables will be found. For instance, we can have a variable "a". It is the key within this associative array, and the value for that variable or that key will be in the value for that key in an associative set, depending on the current state of the program.

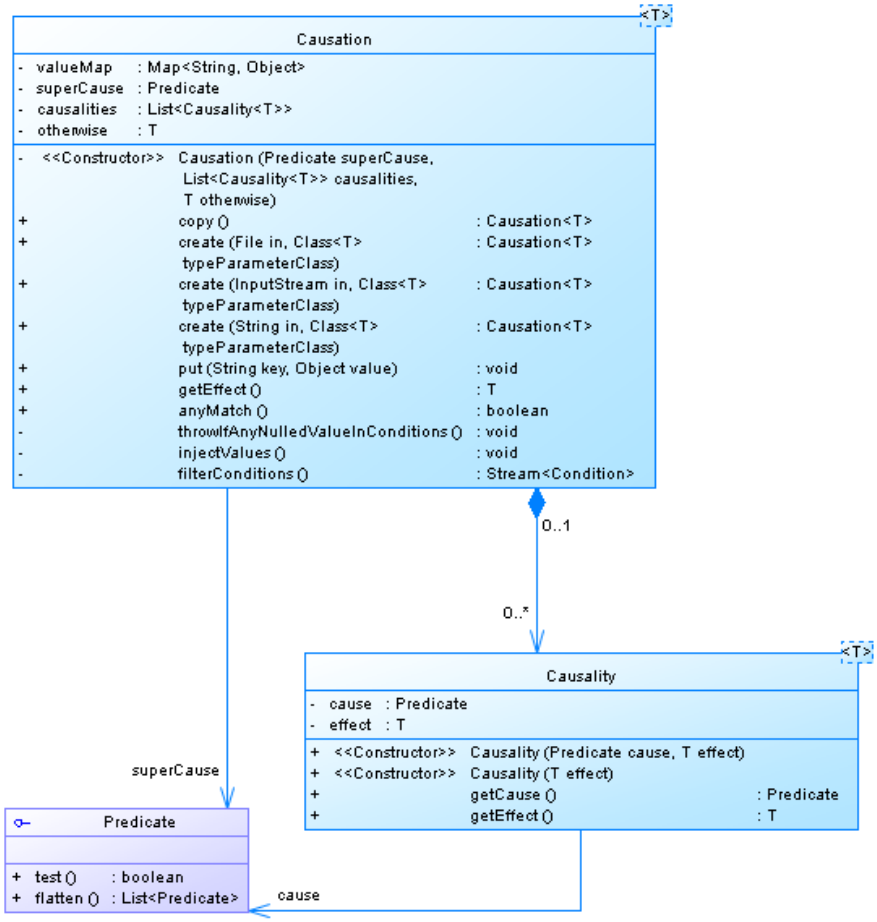


Fig. 5. Hierarchy of Causality classes

```
public T getEffect() {
    injectValues();
    throwIfAnyNulledValueInConditions();
    if (!superCause.test()) {
        return otherwise;
    }
    return causalities.stream().filter(c->c.getCause().test()).findFirst()
            .orElse(new Causality<>(otherwise)).getEffect();
}

private void injectValues() {
    superCause.flattened().flatMap(Predicate::flattened)
            .filter(x-> x instanceof
Condition).map(Condition.class::cast).forEach(c->
c.setV1(valueMap.get(c.getKey())));
    filterConditions().forEach(c-> c.setV1(valueMap.get(c.getKey())));
}
```

Fig. 6. Examples of methods of Causation class

4. Example program

An easy-to-use java program was developed as an example of the utilization of the library. Several model classes will be constructed that will inherit from the Plac (Land) class and will represent residential, commercial or agricultural facilities. There will be some arbitrary variables that will be used to select the appropriate class from the class model.

The arbitrary variables that will be considered are the area, number of rooms and number of floors. Over these random variables, certain predicates will be set up using the JSON format and the Causality library. The effects of the Causality object will be String objects, which will then be passed to the static method create of the Plac class that uses the Factory pattern, which, based on the passed String object, selects the correct class that is a Plac or a subclass of the Plac class.

The first level of selection corresponds to the search for an effect that can be a class of Plac, or type of Building, or type of Field or House type. If the effect of the first level is the type of Building or House type, the program will further bend to the next selection through newly defined rules from some other Causation data.

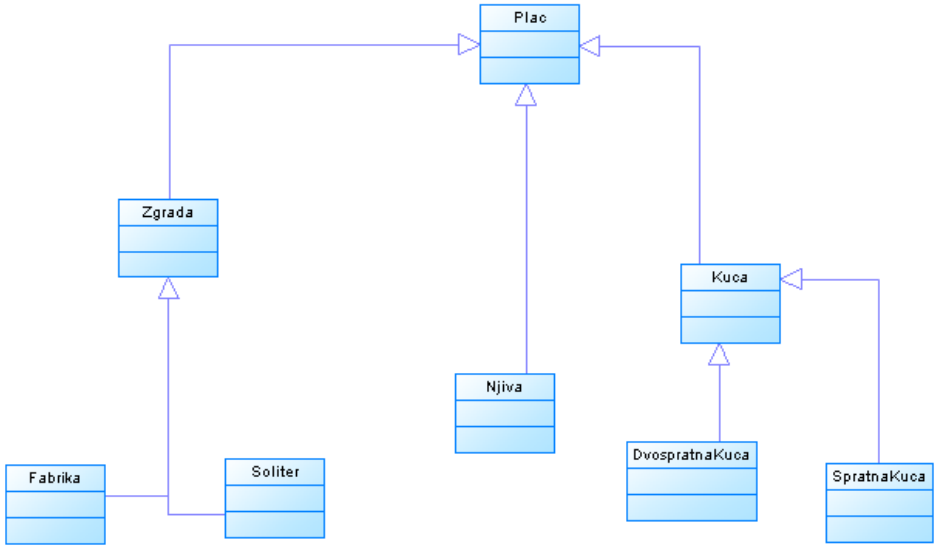


Fig. 7. Hierarchy of Land classes

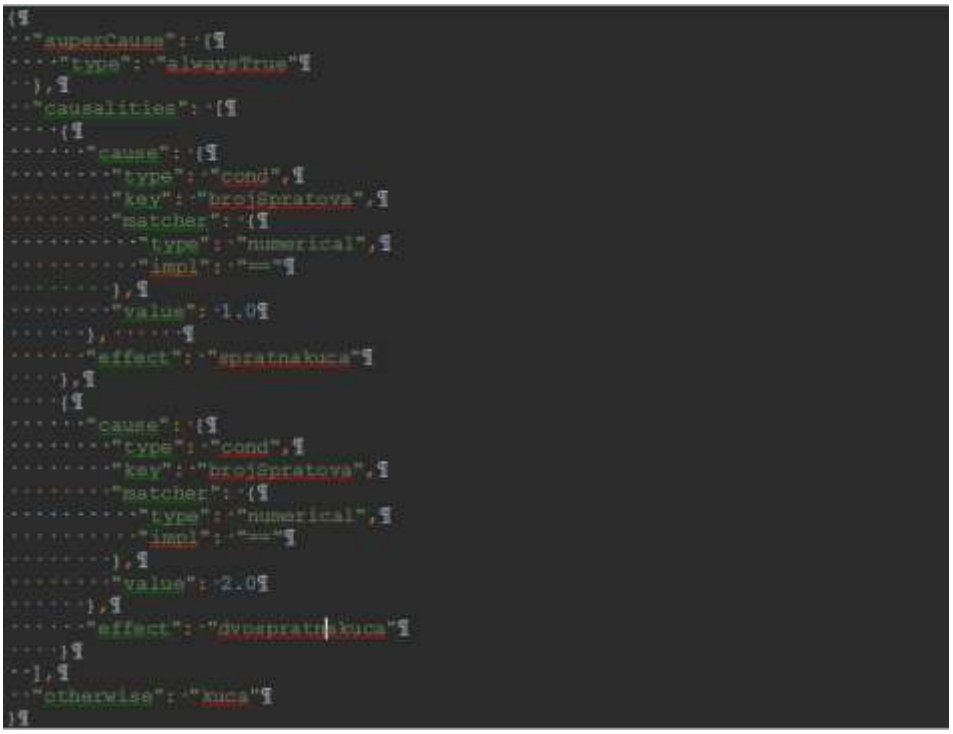


Fig. 8. Example of JSON text used for the example program

```
if (p1 instanceof Kuca) {
    Causation<String> causation2 =
Causation.create(Main.class.getResourceAsStream("/c2.json"), String.class);
    causation2.put("povrsina", povrsina);
    causation2.put("brojSoba", brojSoba);
    causation2.put("brojSpratova", brojSpratova);
    Plac p2 = Plac.create(causation2.getEffect());
}
```

Fig. 9. Part of the code of example program

If the type of object p1 is House, this part of the code is executed. First, the c2.json data is read and parsed as a Causation object typed over the String class. After that, this Causation object, causation2, sets values within its valueMap object with the forwarded arguments in the program. Finally, you can get the effect of the String object type, which is passed to the factory method from the Plac class, and this gives an object that can be a type of House, SpratnaKuca or DvospratnaKuca.

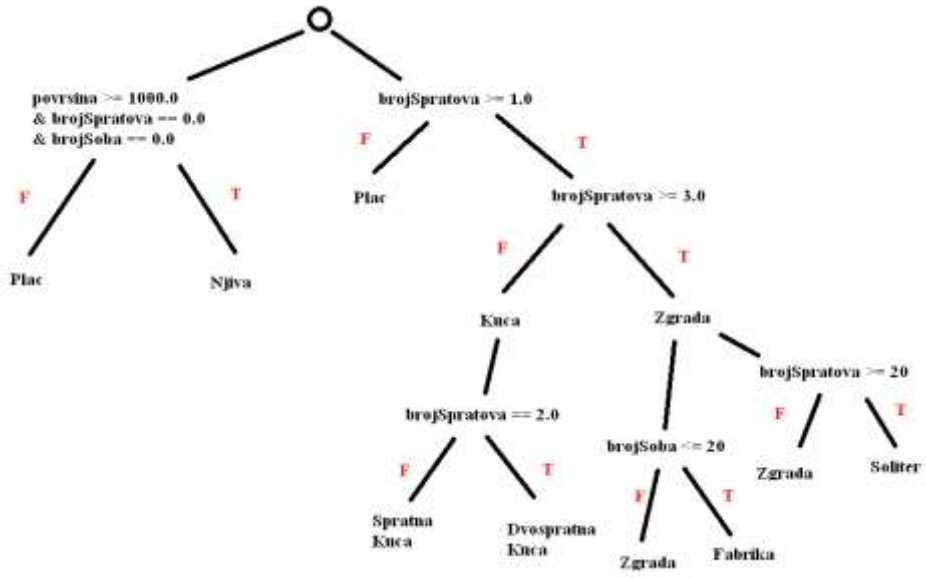


Fig. 10. Conditional tree of the program example

5. Conclusion

Predicate classes are class implementations of Predicate interfaces and they represent the predicate abstraction. Matching classes are so-called "matcher" classes of a type T and they compare two values of type T in a way. The classes of causality are higher-class classes in the library that deal with the already mentioned classes and interfaces and serve for JSON parsing and selecting the right effects across predicates that compare arbitrary variables.

Further improvements to the Causality library could relate to branching support with a single causality class, so that the Causation object could actually consist of several Causation objects and thus divide the effects within a single Causation object meaning that no more Causation objects should be used as shown in the example from the previous chapter.

It is also important to note that this library is dependent on another open-source library, fasterxml-jackson version 2.7.0, which is used to parse JSON text formats. The reason is that this library comes with its JSON parser that is used to determine the Causation object in a text format. Further improvement would be the writing of custom AST parser for a plain formal-predicate mathematical inscription that would contribute to greater readability.

References

1. Gosling, J.; Joy, B.; Steele, G.; Bracha, G.. "The Java Language Specification, 2nd Edition", http://docs.oracle.com/javase/specs/#237601
2. Introducing JSON, http://www.json.org/)
3. https://github.com/FasterXML/jackson-docs
4. Encyclopædia Britannica, Formal system definition, (2007).
5. Sybase, PowerDesigner 16.1 Documentation, (2011).
5. Grady Booch, Object-Oriented Analysis and Design with Applications, (2007)., Addison Wesley Longman Publishing Co., Inc. Redwood City, CA, USA, ISBN:020189551X

Remote Control and Vehicle Management Using ICT

Gordana Jotanović[1], Vladimir Brtka[2], Željko Stojanov[2], Goran Jauševac[1]

[1] Faculty of Transport and Traffic Engineering, Doboj, Republic of Srpska, Bosnia and Hercegovina
gjotanovic@yahoo.com
gjausevac@gmail.com

[2] Technical Faculty "Mihajlo Pupin", Zrenjanin, Republic of Serbia
vbrtka, zeljko [@tfzr.uns.ac.rs]

Abstract: The characteristics of a man's lifestyle in urban time are the use of ICT and "forgetfulness". The idea of this paper is to implement these characteristics in the field of transport and traffic engineering for the safety and reflection of general-purpose vehicles and passengers in it. In the paper we propose the architecture of the system in the function of ICT that addresses the problems of vehicle and passenger safety, the technical condition of vehicle safety, vehicle servicing, validity of driver documents, etc. The mentioned problems are solved by the method of remote control and general-purpose vehicle management using ICT.

Keywords: Remote control, vehicle management, ICT (Information and Communications Technology).

1. Introduction

Today, it is possible to remotely control and management of vehicles or checks the states of vehicles in real time using ICT (Information and Communications Technology). Classic electronic remote device such as starters and alarms can work within a distance close to a vehicle, for example they cannot be used to remotely control a vehicle parked in a parking lot from an office of a tall building or apartment house and also it is difficult to check the state of a vehicle using them [1]. In order to resolve this problem, the present invention presents general-purpose remote control technology that enables a user to control his or her vehicle located at a remote location. These and similar systems allow the driver to operate the vehicle remotely in order to increase the safety of vehicles and passenger in vehicle.

2. Existing Systems for Remote Control and Vehicle Management

Remote control system for controlling a vehicle with a smartphone includes the smartphone programmed with an application that can be installed in the phone via a computer or downloaded from the Internet [2]. The system also includes the vehicle that has been equipped with a radio antenna adapted for communication with the smartphone. Using smart phones, it is possible to remotely control vehicles or check the states of vehicles in real time [2]. Among these ICT device, users to remotely start their vehicles and to check the states of their vehicles using smartphones anytime and anywhere, because the vehicles are equipped with the function of communicating with the smartphone's [1, 2].

One of the negligence type scenario, leaving children in a car for few moments; especially in hot or cooled atmosphere, can cause a catastrophic tragedy to occur. Such systems has been created by developing a smartphone application and a surveillance system connected together to monitor the temperature and the presence of a child inside the car. The system starts to measure temperature inside car via temperature sensor. At the same time, it checks the presence of child inside car via motion sensor. When system detects the presence of a child and the temperature inside car reaches unsafe limit, it will alert the caregiver via smartphone application [3].

Systems for efficient automotive security are implemented for anti-theft using an embedded system occupied with a GPS (Global Positioning System) and a GSM (Global System of Mobile) [4]. Using GPS locator the target current location is determined and sent, along the various parameters received by vehicle's data port, via SMS through GSM networks to a GSM modem that is connected to laptop and/or smartphone [5]. These systems are very safe to report emergency situation and alarms the car driver [6, 7].

The safety of private vehicles, drivers and passengers are a major concern nowadays. There are authentic systems that give alert in cases of unlocked doors and open windows on cars [8]. The system owner can completely control his vehicle using his/her smartphone and some odder ICT [9].

3. Methods

This paper presents a possible architecture that has the task of alarming drivers' faults due to his carelessness and forgetfulness. The idea of paper is to merge several existing systems dealing with the remote control in a unique system that alerts every careless behavior of the driver to the vehicle and passengers.

4. Proposal Architecture of the System for Remote Control and Vehicle Management

System architecture is ICT that are implemented in vehicles, Central locations and client computers or smartphone. All the collected data in vehicles (sensors data) are safely transferred (GPS, GSM, Wireless, etc.) to Central location where they are stored and processed in central server and, if needed, forwarded to other ICT systems. Remote access is available via smartphone or laptop, which enables user to directly connect either to the vehicle, see Fig.1.

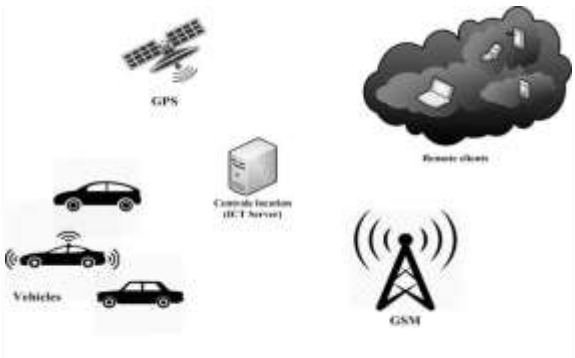


Fig.1. System architecture for remote control and vehicle management.

With standard architecture, the basic parts of the system are:

- Mobile device
- Central location
- Set the alarm application
- AVL (Automatic vehicle location) system

4.1. Mobile device

The devices of this system are intended for collecting data in vehicles. In addition to the position of the vehicle, devices can also collect a range of other information, driver identification, vehicle temperature, door status, key, switch, etc.

Parts that are in the vehicle are:

1. Mobile devices. Mobile devices used for data acquisition and communication with Central location.

2. A wide range of sensors for collecting desired data and connecting to Mobile devices (driver identification, temperature measurement, motion sensor, door status,

keys, switch, etc.). Adapters and sensors needed for required data collection (temperature, windows switch, motion sensor, door switch, etc.).

3. Set the alarm application: It serves to exchange messages and data between the driver and the Central server. Applications for Smartphone, Laptop for message and data exchange.

4.2. Central location

Central server with installed and configured operating system and SQL database, GSM and SMS services used for communications with Mobile devices.

- AVL (Automatic Vehicle Location) gateway for distributing data acquired from Mobile devices to other AVL systems.
- AVL server for receiving data from other AVL systems.
- Xchange server for automatic or semiautomatic data exchange with other ICT systems.
- Administrator application for system services and device configuration, maintaining user rights and system supervision.

4.3. Set the alert alarm application

Set of applications that influence the safety of the vehicle and the control efficacy of the system by the operator. Managing vehicle safety is a challenging job that lasts 24 hours a day, alarming is a key part of vehicle control. Timely information is now crucial to reducing possible damage and removing the consequences caused by the negligence and obscurity of the driver. A set of alarm applications allows the operator to control the vehicles 24 hours a day. The system operator can define the alarm events or use predefined alarms. The system sends alerts via SMS, E-mail. Standard alarm events are driving without driver notification, activation of the sensor in the vehicle (e.g. temperature measurement, door/window status). The system operator selects the way the alarm is distributed and the alarms are distributed. Each vehicle can have different alarms and different people can receive alarms for each vehicle.

4.4. AVL system

AVL (Automatic Vehicle Location) is a system for automatically determining and transmitting the geographic location of a vehicle. This system used with a GPS (Global Positioning System) and a GSM (Global System of Mobile). Proposed system determines the current location and status of vehicles using Google Earth. Using GPS system the target current location of vehicle is determined, along the various parameters received by vehicle's data port, via SMS through GSM network to a user smartphone or laptop.

5. Vehicle Management with ICT

Vehicle management of allows users to easily gain insight into the status of individual vehicles, without having to apply complex dispatch centers and the purchase of expensive ICT devices. Using the Internet or mobile phones, they can have real-time views of the vehicle's positions receive alerts on selected smartphones (to deviate from selected vehicle status parameters), automatically receive alerts on SMS and/or E-mail address and affect a whole range of other functions of importance for the remote control of vehicle management.

Systems based on ICT intended for monitoring vehicles as well as systems for controlling, protecting and managing are not a novelty on the market. They have been present for some time among us and as such are adapted to today's accelerated lifestyle.

6. Conclusions

Modern systems for positioning and remote control of vehicles are based on the computer application and technology based on ICT. Systems achieve positive effects both for drivers and passengers. Experiences in existing systems have shown that all investments in the development and application of such a system are economically paid to both users and service providers in several years of exploitation. Unlike existing robust systems, this system is much cheaper and more flexible. Systems for remote control and management of vehicles, based on ICT, allows the vehicle owner to know at any time what alarming situations inside and outside his vehicle are caused by the driver's negligence or/end forgetfulness and to solve them as quickly and effectively as possible. Relevant system can save money for drivers, enable better security of passengers. The vehicle management system enables faster, safer and easier transportation of passengers, as well as reduction of possible damage to the vehicle. The main objective of this system is to collect all operational data in vehicle, so all data is in one place and available in the shortest time to alarm the driver as result of maintenance of vehicles, vehicle documents and passenger safety.

7. References

1. Hyunkyun, C. H. O. I., & Kim, J. (2017). *U.S. Patent No. 9,644,591*. Washington, DC: U.S.Patent and Trademark Office.
2. Ganz, B. L., Liedblad, B. M., & Thiemann, H. (2017). *U.S. Patent No. 9,569,954*. Washington, DC: U.S. Patent and Trademark Office.
3. Abulkhair, M., Mulla, L., Aldahiri, A., Alkhatabi, H., Alonezi, H., & Razzaq, S. (2017). Sensor Oriented Approach to Prevent Hyperthermia for Children in Car. In *Advances in Human Aspects of Transportation* (pp. 635-640). Springer International Publishing.
4. Maurya, K., Singh, M., & Jain, N. (2012). Real time vehicle tracking system using gsm and gps technology-an anti-theft tracking system. *International Journal of Electronics and Computer Science Engineering. ISSN*, 2277-1956.
5. M. A. A. Khedher, "Hybrid GPS-GSM localization of automobile tracking system," International Journal of Computer Science and Information Technology, vol. 3, no. 6, pp. 75-85, Dec 2011.
6. Rana, T., Shah, A., Rana, P., & Chandak, S. (2017). Smart Vehicle Security, *International Journal of Engineering Science*, *10264*.
7. Ali, J., Nasim, S., Ali, T., Ahmed, N., & un Nabi, S. R. (2009, November). Implementation of GSM based commercial automobile tracker using PIC 18F452 and development of Google Earth embedded monitoring software. In *Research and Development (SCOReD), 2009 IEEE Student Conference on* (pp. 33-36). IEEE.
8. P. P. Wankhade and S. O. Dahad, "Real time vehicle locking and tracking system using GSM and GPS technology-an anti- theft system," International Journal of Technology and Engineering System, vol. 2, no. 3, pp. 272-275, March 2011.
9. Ali Rahnamei, Farnood Khoshnevis, Mina Vajdi, Payam Farhadi, "A Design for CAR Anti-Theft System using Cell Phone", International Journal of Advanced Scientific and Technical Research, [Feb. 2012].