UNIVERSITY OF NATIONAL AND WORLD ECONOMY

INTERNATIONAL CONFERENCE ON APPLICATION OF INFORMATION AND COMMUNICATION TECHNOLOGY AND STATISTICS IN ECONOMY AND EDUCATION

(ICAICTSEE-2012)

OCTOBER 5 – 6TH, 2012

UNIVERSITY OF NATIONAL AND WORLD ECONOMY

SOFIA, BULGARIA

CONFERENCE PROCEEDINGS

UNIVERSITY OF NATIONAL AND WORLD ECONOMY

INTERNATIONAL CONFERENCE ON APPLICATION OF INFORMATION AND COMMUNICATION TECHNOLOGY AND STATISTICS IN ECONOMY AND EDUCATION (ICAICTSEE-2012)

OCTOBER 5 – 6TH, 2012
UNIVERSITY OF NATIONAL AND WORLD ECONOMY
SOFIA, BULGARIA

CONFERENCE PROCEEDINGS

PROCEEDINGS OF THE

INTERNATIONAL CONFERENCE ON APPLICATION OF INFORMATION AND COMMUNICATION TECHNOLOGY AND STATISTICS IN ECONOMY AND EDUCATION

(ICAICTSEE-2012)

EDITOR:

PROF. DR. DIMITER G. VELEV
DEPT. OF INFORMATION TECHNOLOGIES AND COMMUNICATIONS
UNIVERSITY OF NATIONAL AND WORLD ECONOMY
UNSS – STUDENTSKI GRAD
1700 SOFIA, BULGARIA

DVELEV@UNWE.ACAD.BG

ISSUED FOR PUBLICATION: JANUARY 11TH 2013

SIZE: JIS-B5

QUIRES: 39

VOLUME: 150

ISBN 978-954-92247-4-0

ICAICTSEE-2012 PROGRAM COMMITTEES

CONFERENCE CHAIR

Dimiter G. Velev	*University of National and World Economy*	Bulgaria

PROGRAM COMMITTEE

Evangeliy Andronov	*University of National and World Economy*	Bulgaria
César Correa Arias	*Universidad de Guadalajara*	Mexico
Saji Baby	*GEO Environmental Consultation*	Kuwait
Venelin Boshnakov	*University of National and World Economy*	Bulgaria
Donald T. Chang	*Metropolitan State College of Denver*	USA
Sonia Chipeva	*University of National and World Economy*	Bulgaria
Sérgio Manuel Serra da Cruz	*Universidade Federal Rural do Rio de Janeiro*	Brazil
Vladimir Dimitrov	*Sofia State University*	Bulgaria
Cascini Gaetano	*Politecnico di Milano*	Italy
Valentin Goev	*University of National and World Economy*	Bulgaria
Christopher Khoo Soo Guan	*Nanyang Technological University*	Singapore
James K. Ho	*Iniversity of Illinois at Chicago*	USA
Aytekin Isman	*Sakarya University*	Turkey
Sergey Ivanov	*University of the District of Columbia, Washington, DC*	USA
Masayasu Kanno	*Kanagawa University*	Japan
Madjid Malikovich Karimov	*Tashkent University of Information Technologies*	Uzbekistan
Hee Dong Kim	*Hankuk University of Foreign Studies*	S.Korea
Valentin Kisimov	*University of National and World Economy*	Bulgaria
Oleksiy E. Kovalenko	*Institute of Mathematical Machines and Systems Problems, NASU*	Ukraine
Ronald S. Laura	*University of Newcastle*	Australia
Patrick Letouze	*Universidade Federal do Tocantins*	Brazil
Havar Amir oglu Mammadov	*Azerbaijan Technical University*	Azerbaijan
A.K.Mishra	*Motilal Nehru National Institute of Technology*	India
Alexander N. Moiseev	*Tomsk State University*	Russia
Balakrishnan Muniandy	*Universiti Sains Malaysia*	Malaysia
Valentin V. Nechaev	*Moscow State Institute of Radiotechnics, Electronics and Automatics*	Russia

Serghei Ohrimenco	*Laboratory of Information Security, AESM*	Moldova
Veslka Pavlova	*University of National and World Economy*	Bulgaria
David Nadler Prata	*Universidade Federal do Tocantins*	Brazil
A. M. Rawani	*National Institute of Technology*	India
Santhidran Sinnappan	*University Tunku Adbul Rahman*	Malaysia
Kamelia Stefanova	*University of National and World Economy*	Bulgaria
Dimiter G. Velev	*University of National and World Economy*	Bulgaria
A.F. Verlan	*Pukhov Institute for Modelling in Energy Engineering, NASU*	Ukraine
Parviz Ali-Zade	*OKAN University*	Turkey

LOCAL ORGANIZING COMMITTEE

Emil Denchev	*University of National and World Economy*	Bulgaria
Rosen Kirilov	*University of National and World Economy*	Bulgaria
Vania Lazarova	*University of National and World Economy*	Bulgaria
Alexandrina Murgeva	*University of National and World Economy*	Bulgaria
Monika Tzaneva	*University of National and World Economy*	Bulgaria
Ilko Velikov	*University of National and World Economy*	Bulgaria

CONTENTS

Dear Colleagues,

On behalf of the Conference Committees I have the honor and pleasure to thank you all for your true participation in the work of the International Conference on Application of Information and Communication Technology and Statistcis in Economy and Education (ICAICTSEE-2012) which took place on October 5 – 6th, 2012 at the University of National and World Economy (UNWE), Sofia, Bulgaria.

I would like specially to express my gratitude to all of the program committee members for their genuine support without which it would never had happened.

The conference has been dedicated to the 45th anniversary of the establishment of the Department of Information Technologies and Communications at UNWE (1967). The conference has been organized by the Faculty of Applied Informatics and Statistics, UNWE.

Due to the current state of scientific development in all spheres of human activity, the constant knowledge and skill actualization of the academia and researchers in the field of Information and Communicaton Technology (ICT) is an obligatory necessity, especially when the world is in a global economic and financial crisis. The definition of long-term scientific research tasks in this area has a strategic importance, which is even more true for young researchers, lecturers and Ph.D. students.

The conference aims were conducting analysis of the current problems and presenting results of the ICT application in different areas of economy, education and related areas of scientific knowledge; outlining the existing possibilities for the application of modern ICT tools, methods, methodologies and information systems in economy and education; discusing advanced and emerging research trends with a long-term importance in the field of ICT application in economy and education.

The conference has tried to establish an academic environment that fosters the dialogue and exchange of ideas between different levels of academic and research communities.

The conference outcome is 81 published research papers, 24 of which are from foreign participants, the explosion of fresh ideas and the establishment of new professional relations.

I do hope to meet you all again at ICAICTSEE – 2013.

Dimiter G. Velev
ICAICTSEE-2012 Chair

APPLICATION OF INFORMATION AND COMMUNICATION TECHNOLOGIES IN DEVELOPMENT OF E-EDUCATION

H.Mammadov, M.Babanli, Z.Jafarov, A.Suleymanov

Azerbaijan Technical University
babanli@aztu.edu.az

The adoption of the State Program on Informatization of Educational System in the Azerbaijan Republic has accelerated the process of installation of e-university. The primary task within the framework of the program is to intensify the processes in tertiary education, develop human capital and provide a modern level of knowledge transmission technologies.

E-university is a tertiary education management system based on information processing, exchange and dissemination through electronic means. The main objective of its establishment is to facilitate the procedures of providing information services to teaching staff and students, create conditions for correct use of abundance of information, improve quality, and secure transparency. In this regard, the automation of all kinds of information processes in educational institutions is a requirement and integral part of the e-university concept.

E-university includes organization of academic and administrative processes with application of information and communication technologies, as well as software-hardware complex, which allows distance learning and continuous education. This complex is organized by the following systems, which are to operate in a university (corporate) network with a server-client architecture that has a single database:

- *Management-information system*
- *Information provision of the credit system*.
- *Quality monitoring on education services*.
- *Electronic document circulation and e-mail systems*.
- *Information-search systems*.

Management information system ensures automation the following functions within the University:

- Automatic compilation of qualification curriculum (training schedules by course and semester) in accordance with the educational process plan and schedule maintained in database of e-University.
- Students admission announcements, registration of necessary information on admitted and transferred students.
- Automatic compilation of general academic plan in order to reflect each student`s credits in transcripts.
- Automatic compilation of individual academic plan of fresh students.
- Classification of students by credits to be earned.
- Compilation of group plans.
- Preparation of academic plans for new academic year and submission to chairs.
- Automated preparation of curriculums for teachers by chairs.

- Preparation of academic schedule for each group based on semester schedule.
- Automated preparation of online education cards by indicating course hours and topics.
- Registration of students` attendance and earned points in e-journals.
- Automatic calculation of pre-exam points based on notes on e-journals and registration of credits earned by students after e-exams.
- Registration of compositions of commission on defense of graduation theses, protocols and adopted decisions.
- Registration of graduate diplomas.
- Registration of students accommodated in dorm.
- Registration of actual information on personnel.
- Record of staff attendance.

Information provision of the credit system ensures the selection of disciplines and teachers by students, obtaining of information on conditions of earning credits and the access to electronic educational and methodical means:

- Plan and schedule of educational process by specializations.
- Annotation for each discipline and extended program for students prepared by chairs assumed in curriculum of specializations;
- Topics and e-texts of lectures, labs, seminars and training courses
- Wide-screen presentations of lessons.
- Individual work tasks by disciplines.
- Electronic manuals, online lab works.
- Scientific and pedagogical background of teachers.
- Development of individual educational plans of students.
- Selection of updated topics for graduate theses.

Quality monitoring on education services fulfills the scientific-statistical analysis of the educational process on different times and the procedures on control over execution of management functions.

- Control over students credits determined for semester.
- Student attendance monitoring for months and semesters.
- Monitoring of pre-exam scores on disciplines.
- Monitoring of results of exams on disciplines
- Statistical analysis of results of exams.
- Statistical analysis of results of academic year.
- Monitoring of readiness for new academic year by chair, faculty and university.
- Analysis of success and quality indicators taking into account pre-exam points.
- Analysis of teachers' scientific and pedagogical background for participation in competition for holding office and attestation.
- Analysis of scientific and pedagogical activities of staff, chairs, faculties and University.
- Monitoring of educational and methodical provision by specializations.

Electronic document circulation and e-mail systems allow exchange of documents through university and their maintenance via database principle:

- Registration of documents received by University.
- Distribution of orders, instructions and other kinds of documents of the University administration to relevant departments.
- Opportunity to keep track of applications and presentations.
- Registration of documents for students to be provided to official bodies.
- Exchange of information between participants of training.
- Real time collective preparation, coordination and distribution of documents.
- Archiving of video materials on University-scale events, academic seminary, open classes recorded by video cameras installed in relevant rooms.

Information-search systems ensures obtaining data (service, regulatory, analytical, reference, etc.) by interested parties (university board, management, teachers, staff, students, etc) in accordance with their competence.

- Education Law of the Republic of Azerbaijan
- Structural information on University
- Information on Management
- Information on University's external relations
- Information on specialties (specializations), chairs and faculties.
- State Education Standards on specialties.
- Recommendations on Bologna process credit system.
- Regulation on establishment of credit system education in higher educational institutions.
- Regulation on evaluation of knowledge of students of credit system education.
- Regulation on examinations of students of credit system education by disciplines.
- E-library of educational and methodological means prepared by teaching staff of University.
- Archive of scientific articles produced by teaching staff of University.
- Test banks across by disciplines.
- Compositions of Scientific, Scientific-methodic and Dissertation Councils of University.
- Archive of synopsis of defended dissertations.
- Archive of research works, master and graduation theses.
- Current schedule of lessons by students, teachers and classrooms.
- E-inquiry book reflecting data such as URL (website), e-mail, telephone of administration, departments, staff and students.
- Alumni database.
- E-museums reflecting history of development of main areas or objects on each specialization.
- Chronology of significant dates and events on each specialization.

The elaboration of the proposed system starts, first of all, with a two-level investigation of the university as an object of automation:

1. External relations of University is analyzed at macro-level:

- related to the State Committee on Student Admission (list of students)
- Ministry of Education of the Republic of Azerbaijan (graduates, who receive diplomas)
- relations with various domestic and foreign universities
- relations with various ministries and entities.

2. Structural elements and the relationship between them are researched at micro-level:

- e-university infological project (information objects to be maintained in database and their relationship);
- e-university conceptual scheme (list of attributes by main and subsidiary schedules) is developed.

Then the automated database of University is prepared based on relation-modeled database organized with related tables:

1. University Faculties, Chairs (phone, fax, e-mail, URL)
2. Specialty (specialization) and curriculum (discipline annotation, extended program, chair)
3. Provision of e-learning methods by disciplines.
4. Admission plan and admitted plan (featuring information of the State Committee on Student Admission)
5. Teachers (phone, e-mail, URL) and their scientific and pedagogical activity
6. Groups (education level, type, form, department, division)
7. List of topics for graduate these for current year.
8. Auditorium data.
9. Dorm data.
10. Specialty (specialization) training schedules for groups.
11. Delivery of credits earned by students according to individual study plan and e-exams.
12. Academic load of chairs (teaching hours).
13. Schedule of auditoriums.
14. Teaching methods.
15. Teachers curriculum.
16. Electronic journals (attendance and pre-exam point records).

17. Results of e-exams.
18. Students provided with dorm.
19. Presentation of diplomas as a result of decision of Supreme Attestation Commission.

A university-wide corporate computer network is installed. The "Server-client" technology is chosen as the network management method (network operation system). In order to prevent overloading of the network, the sections traffics belong to are isolated.

The database server is installed at a junction of the university-wide corporate computer network.

- Database is stored as files in the network server disc;
- Database management system (DMS) is located in the same server;
- Each using computer of the network has database and interface program (API);
- User applies DMS in the server for receiving or updating information through interface program;
- Structured query language (SQL) is used for applying DMS, i.e. only text of query is submitted by user (customer) to server via network;
- Based on information from physical structure of the database located in server, DMS processes the query and send the outcome to user`s computer.
- Interface program in user`s computer reflect the outcome of the inquiry.

Horizontal portal is established in order to use information resources and services of the e-university complex. As most of the pages of this site are used for internal purposes at university, an internal DNS (web) server is installed taking security into account.

University-wide (corporate) mail server is set up taking into account possibilities for installing a large number of mail boxes, fast and secure document sharing, joint activities, creating reserve copies.

Thus, the installation of the e-university project as part and demand of e-Government will allow the organization of education on the basis of information and communication technologies push the quality improvement of management and educational processes, make them more transparent and attractive, provide the modernization of tertiary education, and positively influence training of specialists.

Functional design of the control of knowledge on base of fuzzy logic

Havar Mammadov[1], Shahnaz Shahbazova[2], Mustafa Babanli[3]

[1] Department of Automation & Control,
[2] Department of Information Technology & Programming,
[3] Department of Metallurgy & Metallography
Azerbaijan Technical University, 25 H.Cavid Ave., Baku, AZ1073 Azerbaijan
rektor@aztu.edu.az, shahbazova@gmail.com, babanli@aztu.edu.az

Abstract: This paper deals with the theoretical and practical bases for the construction of the expert system implemented in a distributed environment. Most of the key issues related to the design of high-quality mathematical elements of the expert system for educational purposes are addressed and offered solutions to. Mathematical formulations of a number of problems of neural networks and expert systems in hybrid subsystem of decisions and assessment are also given.

1.Introduction

One of the most promising ways to improve the functioning of the systems of technical knowledge control is the use of integrated intelligent computer technologies, namely, systems based on heterogeneous knowledge of hybrid expert systems (ES). Hybrid ES present different kinds of knowledge as conceptual, expert, factual, and correspond to different methods of processing [2, 3].

The main task for the development of hybrid systems is the best way to combine different forms of representation and processing methods of knowledge in the decision making process of ES, i.e. the urgent task is to study the opportunity of optimal connection of different mechanisms of knowledge processing to improve the quality, mobility, and efficiency of ES in solving problems and control of knowledge in the state of uncertainty [8,9].

The mobility of ES is due to mobility of knowledge bases (KB), and the possibility of its recharge from different information components (database, database of expertise knowledge (DEK), conceptual knowledge bases (CKB), dynamic files, etc.), as well as a variety of treatments withdrawal. Specification of knowledge in solving problems decomposed them into accurate and inaccurate, complete and incomplete, static and dynamic, single-valued and multi-valued, etc. In addition, the expert knowledge itself is inaccurate because of its subjective nature.[18] The approximation and ambiguity of knowledge leads to the fact that the EC deals with more than one alternative area. Therefore, the incompletion of knowledge allows to use more than one source of knowledge.

2. Process of increase of the structure of the hybrid system

Theoretical construction and software implementation of ANN constantly face the problem of development and complexity of relationships in the general structure of the network, so the system must have a very important new feature - the generation of the structure "from simple to complex". Two terms are used to limit the required capacity - to achieve the required accuracy or by comparison of the Lipschitz constant for the network and a sample of the Lipschitz constant.

2.1. Formulation of the problem

Function F on R is determined by the set of its values at random points in space $(x_1, y_1),...,(x_N, y_N)$. We construct its approximation by combinations $f_i(\overline{X}, \overline{a})$ - from a set of functions Φ that are smooth and continuously differentiable.

Then

$F_1 = F - f(x, a)$ - the error of approximation of the function F;

$F_2 = F_1 - f_1(x, a)$ - the error of approximation of the previous step;

$F_3 = F_2 - f_2(x, a)$

Approximation can be carried out not only by selection of coefficients, but also by choice at each step of functions $f_i(x, a)$ from Φ. Thus, it may be obtained by the expansion of F in a convergent series of the form [16, 17]:

$$F = \sum f_i(a, x)$$

Solution of the approximation can be obtained by minimizing the cost functional corresponding to the square deviation

$$H = \sum (F(\overline{a}^i, \overline{x}^i) - y_i)^2 \to \min$$

The challenge is in the approximation of the function F, given original sample points, using neural network predictor with an unknown number of neurons and the form of the function used in the transmitter of each of the neurons.

The solution can be represented as an iterative process that consists of the following steps:

- Inclusion of a new neuron;
- Optimization of the prediction error of the values in the given points for the current neuron by selecting the function of the transmitter, its parameters and the synaptic weights;

If the desired accuracy is achieved, the process can be stopped, otherwise - the process begins again with the parameters of the already trained neurons are fixed, so that each new neuron is trained to calculate the uncertainty left by the previous ones.

The number of iterations of exhaustion of errors may also be limited by the condition of exceeding the lower bound of the Lipschitz constant for the constructed neural network over the upper bound of the Lipschitz constant sample.

3. Analytical solutions of synapses

Let Y_i be an approximating meaning of another layer. Then Y_0 is the very meaning of the approximated function at the points of the experimental sample, while Y_i and the following are error calculations on the appropriate step.

Education is carried out by optimization of network settings by either of gradient methods throughout the book of problems [1].

Then the training k neuron

$$Y_k = w_{k1}^{(2)} F_k(\alpha_k \sum_{i=1}^{N_{\text{вх}}} w_{ik}^{(1)} x_i),$$

respectively *H* (error function) has the form for the whole book of problems

$$H = \sum_{j=1}^{N_{\text{р}}} \left[w_{k1}^{(2)} F_k(\alpha_k \sum_{i=1}^{N_{\text{вх}}} w_{ik}^{(1)} x_i^j) - Y_k^j \right]^2,$$

that is, as a proximity criterion of the approximated and approximating function, the sum of squared errors for the entire training set is selected.

For training each of the next neuron, the partial derivative functions F_k by the weight of synapses of the first layer $w_{ik}^{(1)}$:

$$\frac{\partial H_k}{\partial w_{ik}^{(1)}} = \sum_{j=1}^{N_{\text{р}}} 2 \left[w_{k1}^{(2)} F_k(\alpha_k \sum_{i=1}^{N_{\text{‰}}} w_{ik}^{(1)} x_i^j) - Y_k^j \right] \left[w_{k1}^{(2)} \alpha_k x_i^j F_k'(\alpha_k \sum_{i=1}^{N_{\text{‰}}} w_{ik}^{(1)} x_i^j) \right],$$

by parameter of neuron α_k

$$\frac{\partial H_k}{\partial \alpha_k} = \sum_{j=1}^{N_{\text{і}}} 2 \left[w_{k1}^{(2)} F_k(\alpha_k \sum_{i=1}^{N_{\text{вв}}} w_{ik}^{(1)} x_i^j) - Y_k^j \right] \left[w_{k1}^{(2)} (\sum_{i=1}^{N_{\text{вв}}} w_{ik}^{(1)} x_i^j) F_k'(\alpha_k \sum_{i=1}^{N_{\text{вв}}} w_{ik}^{(1)} x_i^j) \right]$$

and the weight of the synapse of the second (output) layer $w_{k1}^{(2)}$ corresponding to the given neuron

$$\frac{\partial Y_k}{\partial w_{k1}^{(2)}} = \sum_{j=1}^{N_{np}} 2 \left[w_{k1}^{(2)} F_k(\alpha_k \sum_{i=1}^{N_{\text{вх}}} w_{ik}^{(1)} x_i^j) - Y_k^j \right] \left[F_k(\alpha_k \sum_{i=1}^{N_{\text{вх}}} w_{ik}^{(1)} x_i^j) \right],$$

where N_{np} is the number of examples of the training set.

However, if the calculation of the function *H* is related to the cost of order processing time T_H, the calculation of its gradient in the traditional way takes time of order

$$T_{\text{grad}H} = nT_H,$$

where *n* is the number of variables of the function *H*. Given that the tasks, which are traditionally applied a neural network, the value of n can reach several thousand, an analytical solution for the calculation of the gradient of the error function should be declared inadmissible [4,5,6,].

However, in describing the decisive function F as a network of automatic computation of the gradient of the error function H can be represented as a function of the dual source. In this approach,

$$T_{gradH} \approx CT_H ,$$

where *C* is a constant independent on the dimension n and in most cases is approximately equal to 3 [7].

Thus, we come to write the original problem in the ideology of neural networks.

4. Application of fuzzy logic

Application of fuzzy logic of hybrid ES control knowledge can have at least three implementations: 1) processing fuzzy expert statements, that is, when the premise has fuzzy variables, and inference engine - a mechanism to retrieve data from them, and 2) the use of a matrix of fuzzy relations determining a number of factors and a lot of assumptions. The matrix contains fuzzy variables ratio, a measure of which is represented as a real number [0, 1], and to determine the cause of the state, the transformation matrix and the equations of the form factors of fuzzy relations is generated, and then the resulting system was solved by the composition of the minimum-maximum, and 3) the use of fuzzy conclusions. This approach is most often used in the construction of fuzzy knowledge bases [10].

Using Fuzzy Hybrid ES to solve problems and control parameters of knowledge extends this class of intelligent systems, increases flexibility and mobility, allows in equal computing resources of computers to conduct peer review more options, increasing the reliability and accuracy of the results.

This paper discusses the basic principles of neuro hybrid ES with diverse knowledge and analyzes its performance in the face of uncertainty of parameters of the control object (knowledge), applying as a dynamic knowledge base of the combined models HC [11, 12].

In a hybrid neuro ES, the reference model (RM) is stored in the knowledge base and is refined in the process of acquiring new knowledge. The real model is formed in the database environment, and communication with the EM is conducted via user input. Solution for creating intelligent quality control system of knowledge based on hybrid ES was performed with consideration of the environment of EC.

The hybrid ES consists of the following functional parts: a database that stores reference and factological data of the process, the results of their comparison, conceptual,

infological and physical models of knowledge, knowledge base (KB): static (knowledge stored in the form of expert knowledge (products)), as well as formulas, facts, relationships, tables and concepts of specific subject area); dynamic (knowledge stored combined HC as a reference dynamic processes with partial or complete uncertainty control parameters); inference engine, based on the algorithm of generating the cause-investigation networking events of functional - structural model; the mechanism of adaptation that coordinates the database (DB) and knowledge base in inference depending on the situation; the mechanism of explanation, which is an interpretation process of logical inference; planner that coordinates the process of solving the problem; solver that allows to find effective solutions of direct, inverse and mixed formulations of the problems.[13]

Content, form and algorithms of presentation of hybrid ES have the possibility to vary depending on the complexity of the simulated situation, the specific and individual features of the user.

The expert-user presents expert knowledge in the form of sets of examples. Internal presentation of expert knowledge is the tree of output. A set of examples is described with attributes and contains examples of the same structure, defined by its attributes, which can be linked by logical transitions. In this case, the relevant output trees get combined in such a way that at the terminal nodes of one tree adds another tree [14].

Computational Model ES and DB for solving problems in the face of uncertainty is given in summary form: W = <A, D, B, F, H>, where A is a set of attributes of database and knowledge base; D - domain (meanings of attributes database and knowledge base); B - a set of functional dependencies defined over the attributes; F - many descriptions of all types used in the B functional dependencies; H - set of fuzzy relations over a set of attributes A.

Please note: each hybrid ES has its specific requirements for the form of knowledge representation, and since they are different (frames, semantic networks, database, concepts in the knowledge base of ES, neural networks, fuzzy logic, genetic algorithms), then, even within the framework of a single information Space of hybrid ES, to combine diverse knowledge is quite complicated. For example, in the hybrid ES heterogeneous knowledge is stored in the static ES, while dynamic knowledge about the state of knowledge in neural networks [15].

Modern information computer technologies (based on the approach of OLE-technology) make it easy to share diverse knowledge within a single information space of hybrid neuro ES. In conclusion, it should be noted that the foregoing in the approach to the construction of intelligent system control and knowledge-based hybrid ES functioning in the face of uncertainty, you can:

• Actively apply diverse knowledge (conceptual, structural, procedural, factual, rule base with membership functions, rules, and unclear rules database, knowledge base, BEK, procedures) with a combination of mechanisms for effective output solution for the knowledge of the student;

• synthesize and refine the conceptual model is a representation of heterogeneous knowledge among relational database (Access, FoxPro and Informix), managed databases and interacting with the kernel hybrid ES;

• efficiently solve the optimization problem and the distribution of information flows for individual subsystems ES with heterogeneous knowledge under uncertainty[19].

The technique of construction of heterogeneous knowledge hybrid neuro ES to monitor the student's knowledge in the face of uncertainty is the following:

The formalization of the domain (the development of a conceptual model).

Description of the model of knowledge as separate concepts (knowledge) in KB.

BR formation with a base of rules as the control components intellectual core.

Description of heterogeneous data to monitor the student's knowledge in the individual subsystems Hybrid Expert System (DB, KB, an expert knowledge base, a graphical database design files).

Model selection neuro nets and learning rules.

Development of the software part of fuzzy logic.

Distribution of information flows between the ES and its individual subsystems.

Testing of individual subsystems with heterogeneous ES knowledge.

Testing hybrid neoro fuzzy ES.

Conclusions

Mathematical models of constructing expert systems implemented in an intranet environment of the university and its uses in the training process were validated. Most of the key issues related to the design of the university Intranet and the expert system for educational purposes are addressed and offered solutions to.

Modeling key aspects of the main components of the complex obtained the possibility of integration of elements of the expert system in the learning process, the possibility of accumulation and database development, and building a knowledge base built on powerful and flexible rules.

The methods were proposed to simulate ANN of optimal complexity, which will improve the efficiency of decision-making power, without having to re-design and mathematical modeling to fine-tune the system.

References

[1] Bellman R., Zadeh LA, Decision-making in ambiguous circumstances, issues analysis and decision-making, Springer, 1976, pp. 180-199.

[2] Bernshteyn L.S., Bojenyuk A.V. Fuzzy models of decision making: deduction, induction, analogy. Taganrog, Univ Tsure, 2001. pp. 78-99.

[3] Bouchon-Meunier B., Yager RR, Fuzzy Logic and Soft Computing (Advances in Fuzzy Systems: Application and Theory), World Scientific, 1995, pp. 84-93, 103-119.

[4] Hanss M., Applied Fuzzy Arithmetic: An Introduction with Engineering Applications Springer, 1 edition, 2004, pp. 100-116, 139-147.

[5] Heaton J., Introduction to Neural Networks for C #, 2nd Edition, Heaton Research, Inc.; 2 edition, 2008

[6] Nikravesh M., Aminzadeh F., Zadeh LA, Soft Computing and Intelligent data analysis in oil exploration, 2003, pp. 273-287.

[7] Nikravesh M., Zadeh LA, Kacprzyk J., Soft Computing for Information Processing and Analysis, 2005, pp. 93-99.

[8] Shahbazova Sh., Freisleben B., A Network-Based Intellectual Information System for Learning and Testing, Fourth International Conference on Application of Fuzzy Systems and Soft Computing, Siegen, Germany, 2000, pp. 308-313.

[9] Shahbazova Sh., Zeynalova S., Decision-Making In Definition Of Knowledge In The Conditions Of Uncertainty Of Educational Process, PCI-2010, Volume I, "Elm", 2010, pp. 305-310.

[10] Zadeh LA, Kacprzyk J., Fuzzy Logic for the Management of Uncertainty, Wiley-Interscience; First Printing edition, 1992, pp. 75-84.
[11] Zadeh LA, Klir GJ, Yuan B., Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A. Zadeh, 1996, pp. 60-69.
[12] Zadeh L.A. The new approach to the analysis of complex systems and decision processes, Mathematics Today, Knowledge, 1974, pp. 23-37
[13] Zadeh L.A. The concept of linguistic variable and its application to the adoption of approximate solutions, Springer, 1976, pp. 140-164.
[14] Badami, Chiang, Khedkar, Marcelle, Schutten, "Industrial Applications of Fuzzy Logic at General Electric", Bonissone, Proceedings of the IEEE, 83(3), pp. 450-465
[15] Jang and C.-T. Sun, "Neuro-Fuzzy Modeling and Control", J.S.R. Proceedings of the IEEE, 83(3), pp. 378-406
[16] Laurene V. F., Fundamentals of Neural Networks: Architectures, Algorithms And Applications, Prentice Hall; US edition, 1993, pp. 103-121.
[17] The Fuzzy Logic Toolbox for use with MATLAB, J.S.R. Jang and N. Gulley, Natick, MA: The MathWorks Inc., 1995
[18] Yager R., Filev D. Essentials of fuzzy modeling and control / New York John Wiley and Sons, 1994.
[19] Zadeh L.A., A new approach to the analysis of difficulty systems and decision processes, Mathematics Today, Knowledge, 1974, pp. 23-37

Reporting a Code Refactoring and Evolution of an Academic Project Management Web System

Fernando Chagas, Glaubos Climaco, Mariwaldo G. Caetano, Ary H. M. de Oliveira, and Patrick Letouze

Department of Computer Science, Federal University of Tocantins, Brazil
patrick.letouze@gmail.com, fernando.chagas@live.com, glaubos_7@hotmail.com

Abstract. This paper reports a code refactoring experience in an undergraduate level of an Academic Project Management web-based system. The system is called SIGD and it is based on project management concepts to support the administration of a postgraduate program, though it can be used at an undergraduate level. We have used the Evolutionary Acquisition Interdisciplinary Research Project Management (EA-IRPM) strategy together with scrum and the MVC pattern. Our main result was a completely refactored code for the new version of the system. Besides that the system's code has become easier to maintain and to develop as expected. Finally, the students report that they have increased their development skills in a professional level, which signals that a specific discipline of code refactoring could be used as a way to help students to mature professionally.

Keywords: Agile development, experience report, project management, scrum, refactoring, web system.

1 Introduction

Previously, we have developed an academic project management web system called SIGD [1]. We conceived it based on project management concepts to serve as a tool for improving results on a postgraduate program. It's started in a classroom where we simulated a software house using role-playing and a problem-based learning approach [1]. However, we had some productions problems such as unreliable code, small scale development and adaptation issues for future changes, as a consequence of the student's inexperience, who were not much acquainted with programming frameworks and agile development techniques. After all, our objective was to prototype the system, and most importantly, to provide students a real software development experience in a team environment as close to reality as possible.

In order to facilitate improvements and the system's evolution, we decided that a code refactoring should take place. This work reports our experience applying it in our system. We employed agile development techniques and tools such as Scrum, Hibernate, Spring: security framework, Pretty Faces and Maven. Moreover, we used the Evolutionary Acquisition Interdisciplinary Research Project Management methodology (EA-IRPM) [2], which is the strategy employed by NDS – the Software Development Nucleus of the Federal University of Tocantins, where the work was done.

About the new version, in refactoring the older one, we confirmed that production time greatly reduces by applying agile development techniques and tools. An especial improvement was scalability providing a better basis for future developments.

We organized this paper in accordance to the IMRAD structure: introduction, methods, results and discussion; which is adopted as part of the Uniform Requirements for Manuscripts Submitted to Biomedical Journals of the International Committee of Medical Journals Editors, 2008 update. We believe that adopting this structure will help search engines in international databases to store and to retrieve information within research papers facilitating meta-analyses and systematic reviews.

2 Methods

In this section we describe the EA-IRPM [2] – subsection 2.1, the development methodology that we used, which was also used in [3], [4]. Then we describe the tools and techniques that we had employed – subsection 2.2.

2.1 Evolutionary Acquisition IRPM

The Interdisciplinary Research Project Management – IRPM [5] is an approach for interdisciplinary research of real problems using Project Management concepts [6] and problem-based learning [7], [8]. IRPM's schematic is present in Figure 1, but first let us briefly review the Project Management phases: (1) Initiation: to determine project goals, deliverables and process outputs, to document project constraints and assumptions, to define strategy, to identify performance criteria, to determine resource requirements, to define the budget and to produce a formal documentation; (2) Planning: to refine project, to create a work breakdown structure, to develop the resource management plan, to refine time and cost estimates, to establish project controls, to develop the project plan and to obtain the plan approval; (3) Execution: to commit resources, to implement resources, to manage progress, to communicate progress and to implement quality assurance procedures; (4) Control: to measure performance, to refine control limits, to take corrective action, to evaluate effectiveness of corrective action, to ensure plan compliance, to reassess control plans, to respond to risk event triggers and to monitor project activity; (5) Closing: to obtain acceptance of deliverables, to document lessons learned, to facilitate closure, to preserve product records and tools, and to release resources.

In IRPM, Initiation phase begins with choosing the real problem to solve and identifying at least two fields for an interdisciplinary approach. These fields are necessary to: document the real problem constraints and assumptions; define strategy; identify performance criteria; determine resource requirements; define budget; and produce formal documentation. Planning phase consists of refining project and analysing the real problem through studying the chose fields. These studies may produce a new fundamental or methodology. Then in Execution phase, even if new concepts are not obtained, an educational material may be prepared and used in class for a problem-based learning approach, or else the new technology may be implemented and applied. Moreover, if in Planning phase controls were established then educational, technological, economics and social parameters may be available for measurement, allowing Control phase to be performed. Finally, after analysing measurements, papers should be written as part of Closing phase.

The incorporation of EA into IRPM is presented in Figure 3 [2]. It shows that it is inserted into phases Planning, Executing and Control. In Figure 3, RA means Requirements Analysis of: (1) general for the system and specific for the core; and (2) user feedback,

technological opportunities and evolving threat. Hence, in Planning phase the attempt to develop a new fundamental or methodology consists of generating a preliminary system architecture beginning with RA 1, and then elaborating the concept of operations, and when available, also considering RA 2. Executing phase consists of implementing the core from the preliminary system architecture followed by new definitions and developments of operational tests. Afterwards, the system is applied in a real life situation. Control phase is about refining and updating requirements, which implies in evaluating technology, measuring economic and social parameters, and verifying users' feedback, technological opportunities and evolving threats, that is, RA 2.

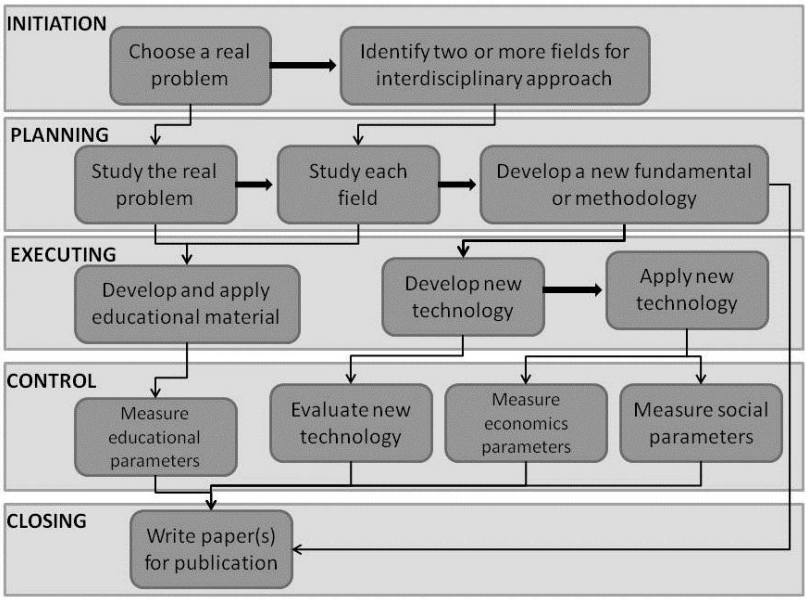


Fig. 1. The Interdisciplinary Research Project Management diagram.

Evolutionary Acquisition (EA) starts with the requirements analysis [9], Fig.2. After defining the "general" requirements for the system and the "specific" requirements for the core, the concept of operations is elaborated. Then together with a requirements analysis of user feedback, technological opportunities and threats evaluation, the preliminary system architecture is developed. From the system architecture a core is produced. New definitions and developments with an operational test may result in a new version of the core. Then with experience and use, new requirements refinements and updates are identified and used to improve or develop a new core.

Additionally, Evolutionary Acquisition separates the core of the system into blocks. A particular block can have several releases, including a final one. If the system is a software, then software engineering techniques may be applied. This paper proposes the use of MVC as the software architecture for the Interdisciplinary Research web service system.

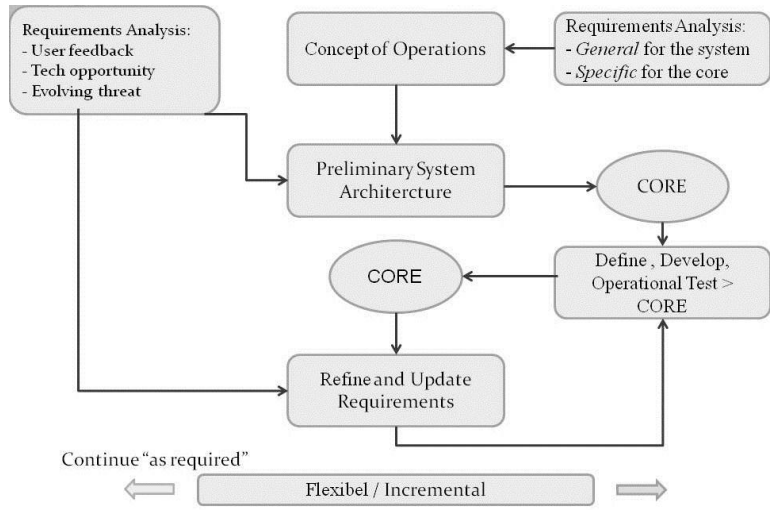


Fig. 2. The Evolutionary Acquisition model.

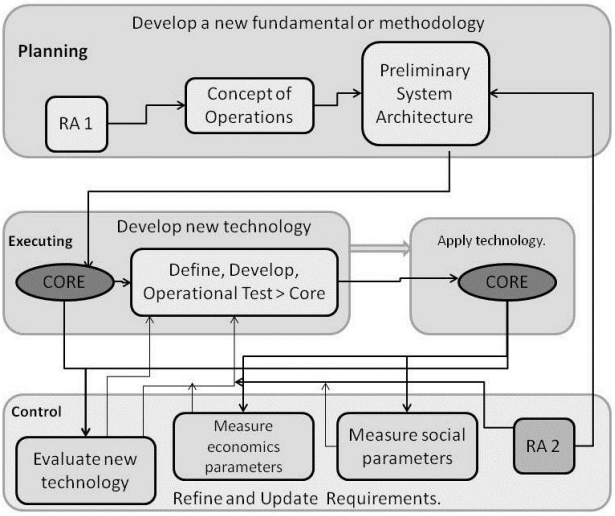


Fig. 3. The insertion of Evolutionary Acquisition into IRPM

2.2 Tools and Techniques

We had used the Model-View-Controller (MVC) pattern as the software architecture [10], [11]. It intends to separate the business logic – model, the user interface – view, and the user input – controller. As a consequence, it provides a way to split features to independent development, testing and maintenance. Basically, the model represents the application data and the business rules that command data access and its modification. It also keeps business

state and provides to the controller the ability to access encapsulated features. The view displays the system's state and the controller sets the application behavior. Figure 4 presents the MVC.

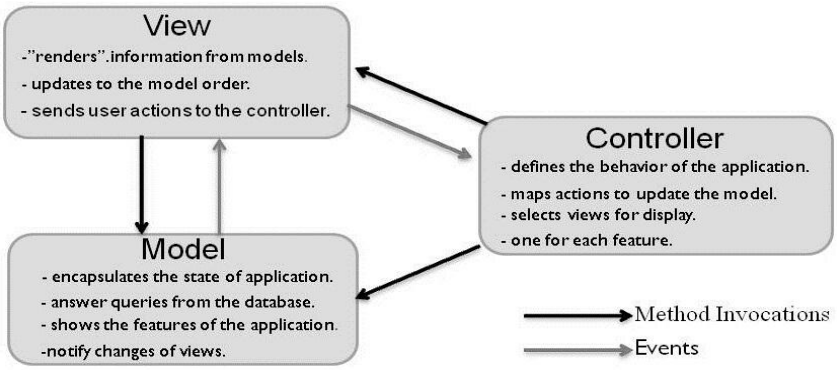


Fig. 4. Model-Viewer-Controller pattern.

We had also used: Java as the object-oriented programming language; Javaserver Faces as the standard for building user interface [12]; Prime Faces as the source Javaserver Faces components [13]; Glassfish as the application server [14]; Netbeans IDE as the integrated development environment (IDE) [15].

Then differently from the first version, we had used: Scrum as a process of iterative and incremental development for an agile software development [16]; Hibernate as an object-relational mapping (ORM) library for the Java language [17]; Spring: security framework 3 as a framework for access control and authentication [18]; Maven as a tool for management and automation of projects in Java [19]; Pretty faces as a filter-based servelets extension with support to JSF and to create URLs [20]; MySQL as the relational database management system [21].

2.3 Refactoring Methodology

An example of refactoring methodology was presented in [22]. Basically, it is divided in two phases: preparation and search. The preparation phase consists of refactoring motives, that is, to extract reusable components, to loose coupling between components, to improve design. While search phase is to: select specifics parts of the project to do the refactoring; propose new aspect-oriented program refactoring - the selection is done around the applicability and situation; determine what refactoring activities can be applied; set priority for refactoring; apply refactoring patterns, which implies in testing cases and realizing if it did not broke anything; and analyze how the refactoring is affecting the software quality.

We had used the example as a starting point, but our system's documentation was incomplete and its source code was poorly comprehensible, which had led to a complete source code refactoring decision. Therefore, our motivation was to create a system that had a complete documentation, high source code comprehensibility, all aiming to increase its life

cycle. As a consequence, we had started with a design refactoring, which had considered aspects refactoring. Then we had set priorities followed by case testing for finally analyzing quality improvements like simplification, code length and reusability.

3 Results

In its first version SIGD was an educational experience with two phases. The first was composed by a team of 20 students using Unified Process and they had freedom in the process; and the second phase was composed by 2 students.

Our second version was developed by two students using Scrum. The version 1.0 had proposed a system, unrestricted file uploads, initial module of internationalization, and the user may have a class in which he was registered. In contrast, the second version included the project area, chapters, upload restricted and validated, hierarchical changes in the relationship of the objects optimizing the search, automatic generation of projects and chapters, banking functionality, internationalization, a user now has several classes avoiding the creation of multiple entries if the system was tapped by another discipline, help manual, revision of heuristics, decreased response time and dynamic pages rendering via Primefaces' ajax.

In table 1, we present a comparison between versions 1.0 and 2.0. In relation to the frameworks used in version 2.0, as hibernate, it transforms the classes into tables; it releases the developer of manual work in addition to maintain the program portable to any database. Fig. 5 is an example that depicts the apparent reduction in the code using hibernate. Spring Security brings a complete solution for all kinds of security needs, and it automates security removing manual work which may be dangerous practice. About Maven, it is a simple basic project structure to code compilation and to run unit tests, and to build the jar file extension. For last, the URLs have become "friendly" with Pretty Faces.

Table 1**.** Comparison between release 1 and 2.

Tool	**SIGD 1.0**	**SIGD 2.0**
Software architecture	*MVC*	*MVC*
Programming language	*Java*	*Java*
Server	*Javaserver Faces*	*Javaserver Faces*
Component suite	*Prime Faces*	*Prime Faces*
Application server	*Glassfish*	*Glassfish*
IDE	*NetBeans*	*NetBeans*
Development method	*Unified Process*	*Scrum*
Object-relational	-	*Hibernate*

mapping		
Application framework	-	*Spring: security framework 3*
Build automation tool	-	*Maven*
URL rewrite filter	-	*Pretty Faces*
Database	*PostgreSQL*	*MySQL, PostgreSQL*

Versão 1.0

```
public void insert(Curso curso) throws ExceptionDatabase {
   Conexao con = new Conexao();
   try
   {
      con.open(false);
      String sql = "select nextval('sigd.idcurso_seq')";
      ResultSet rsSeq = con.execQuery(sql);
      if (rsSeq.next()) {
         curso.setIdcurso(rsSeq.getInt(1));
      }
      sql = "INSERT INTO sigd.curso(id_curso, codigo_curso, nome_curso, total_horas) VALUES (?, ?, ?, ?)";
      PreparedStatement pst = con.getPreparedStatement(sql);
      pst.setInt(1,curso.getIdcurso());
      pst.setString(2,curso.getCodigo());
      pst.setString(3,curso.getNome());
      pst.setInt(4,curso.getTotalHoras());
      con.qryExecute(pst);
      con.commit();
   }
   catch (SQLException e) {
      con.rollBack();
      throw new ExceptionDatabase(e.getMessage());
   }
   catch(NullPointerException e){
      con.rollBack();
      throw new ExceptionDatabase(e.getMessage());
   }
   finally {
      con.close();
   }
}
```

Versão 2.0

```
public void insert(Class beInserted) {
    this.entityManager.persist(beInserted);
    this.entityManager.flush();
}
```

Fig. 5. Code simplification.

4 Discussion

Despite the time required for refactoring the system, the final results compensated the loss of time, because its code is optimized and it is ready for more updates and new developments. Allied to this, its evolution has become much faster due to the use of tools and agile development methods.

The use of more frameworks and efficient production methods made possible to create a more solid platform that will serve us as a basis to other applications, faster coding, and more comprehensible and clean programs with lower failure probability.

Finally, in the new version, we had implemented it with MySQL to test the use of Hibernate, that is, it was enough to only change a few parameters to the system work with PostgreSQL, which have confirmed the database abstraction. Besides all that, the students report that they feel that they had achieved a more professional maturity level, which indicates that a discipline of code refactoring could be used to obtain that for an average student.

5 References

1. P. Letouze, R. A. Ronzani and A. H. M. Oliveira, An Academic Project Management Web System Developed through a Software House Simulation in a Classroom, *IPEDR*, 2011, vol. 14: 587-592.
2. P. Letouze, Incorporating Evolutionary Acquisition into Interdisciplinary Research Project Management. *IPEDR*, 2012, vol. 30: 338-342.
3. G. V. Barbosa, H. D. Soares, G. L. R. de Brito, Sergio M. S. da Cruz, and P. Letouze, "A Social Network to Provide Free Internet Access for Public Schools' Communities". *IPEDR*, 2012, vol. 30: 248-253.
4. P. Letouze, S. A.Naimayer, F. S. Nunes, G. V. Barbosa, A. H. M. de Oliveira, J. Y. Ishihara, and A. Burls, A Medicine 2.0 Architecture: managing transplantation patients. *IEEE Trans. Inf. Technol. Biomed*. To be submitted.
5. P. Letouze, Interdisciplinary Research Project Management. *IPEDR*, 2011, vol. 14, 338-342.
6. K. Heldman. PMP Project Management Professional Exam Study Guide. Sybex, 2009.
7. J. R. Savery. Overview of Problem-Based Learning: Definitions and Distinctions. *IJPBL*. 2006, 1 (1): 9-20.
8. O. Pierrakos, A. Zilberberg and R. Anderson. Understanding Undergraduate Research Experiences through the Lens of Problem-based Learning: Implications for Curriculum Translation. *IJPBL*. 2010, **4** (2): 35-62.
9. "Systems Engineering Fundamentals," Supplement 2-B "Evolutionary Acquisition Considerations," The Defense Acquisition University Press, Fort Belvoir, Virginia, Jan 2001.
10. T. Reenskaug, Thing-Model-View-Editor – an example from a planning system. *Technical note, Xerox PARC*, May 1979. A scanned version on http://heim.ifi.uio.no/~trygver/mvc/index.html
11. T. Reenskaug, Models-Views-Controller. *Technical note, Xerox PARC*, December 1979. A scanned version on http://heim.ifi.uio.no/~trygver/mvc/index.html
12. James Holmes , Chris Schalk, JavaServer Faces: The Complete Reference, McGraw-Hill, Inc., New York, NY, 2006, page 3
13. http://www.primefaces.org/
14. http://glassfish.java.net/
15. http://netbeans.org/
16. Schwaber, Ken. 2004. Agile Project Management with Scrum. Redmond, WA. Microsoft Press. (ISBN 0-7356-1993-X)
17. Will Iverson, Hibernate: A J2EE™ Developer's Guide, 2004, Chapter 1. Overview, Whats is hibernate?
18. Peter Mularien, Spring Security 3, Packt Publishing, 2010, page 18
19. http://maven.apache.org/
20. http://ocpsoft.org/prettyfaces/
21. http://www.mysql.com/
22. S. A. M. Rizvi, Z. Khanam, "A methodology for refactoring legacy code," *Electronics Computer Technology (ICECT), 2011 3rd International Conference on* , vol.6, pp.198-200.

Organization of an integrated system of information security in information and communication systems

A.F. Verlan, M.M. Karimov, S.K. Ganiev, K.A. Tashev

Tashkent university of information technology, str. A.Temur, 108, 100202, Tashkent, Uzbekistan,
dr.mmkarimov@rambler.ru, k.akhmatovich@gmail.com

Abstract. This article describes how to build an integrated system of information security. The structure of an integrated system of information security in information and communication systems are offered.. It also examined the issues of building subsystem identification unauthorized influences. We propose a model for intrusion detection subsystem infocommunication system based on the method of monitoring the state of objects and subjects. This model differs from that used in the model can detect previously unidentified attacks.

Key words: information and communication system, security information, methods detection of attacks, firewalls, management means.

1. Introduction

At the present time, there are two approaches to the creation of information security systems. The first approach involves the implementation of measures designed to prevent unauthorized intruder effects on the scope of the information network. In this approach, the structure of the system include firewalls, access control and security analysis. The second approach is based on the detection of unauthorized entry of the fact the perpetrator, as well as the localization of the place and establish the source of exposure. The basis of the structure of the system at the same time make detection of unauthorized actions that could attempt to fix the breach of information security at an early stage of their development in real time [1, 2]. In these approaches, there are no tools that could simultaneously combine all the characteristics of information security systems [3, 4].

2. Integrated system of information security

To create an integrated security system with all the functional characteristics of the above approaches, it is advisable to include in the structure of the following components: firewall, a means of detection of unauthorized actions, security analysis tools, specialized software and hardware protection (SSHP) and the control unit (Fig. 1).

Firewalls are installed at the point of contact between local and corporate data networks with the networks of public data. They restrict the flow of information coming from the public data network using the internal data network.

Detection of unauthorized actions allows you to monitor and analyze network performance data, as well as attempts to commit violations of network security. Having determined the fact of attack detection system tries to locate the source of the attack and tells the administrator about the need for action. Detection of an unauthorized action is implemented in two modes:

- detection of attacks in the initial phase of their development;
- detection of the effects of impacts on the offending area information network.

There are two types of detection: detection based on host and network-based data. In the first type of media source of information for analyzing the current state is the audit logs. In the second type detection means is performed based on the analysis of network traffic. High effect is achieved by combining both types of detection equipment.

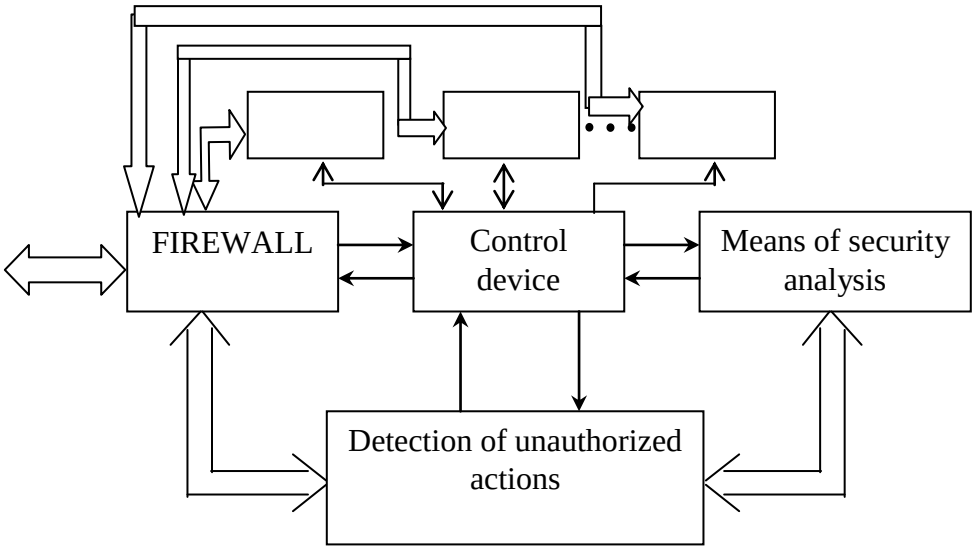


Figure 1.The basic structure of an integrated system of information protection.

Means of security analysis designed to determine the scope of the vulnerability of information networks. They carry out inspection of the network on the availability of vulnerability, and then on the basis of information obtained suggest possible solutions to address them. These tasks can be performed using two mechanisms: passive and active.

Specialized hardware and software (SSHP) protect network data. Application SPAS allows aggregate data, process the trivial situation and notify the administrator of the extraordinary events that require immediate attention to itself.

Control means is responsible for monitoring and process control operation of firewalls, security analysis tools and detection of unauthorized actions.

Operation of the integrated system of information security should be performed only under the supervision of the administrator. Before installing the network to a specific host should be evaluated by the level of protection with the latest security features. Just making sure that the host has no points of vulnerability, an administrator can perform the installation procedure. After completing the installation process with an integrated network information security administrator needs to perform: scanning the information sphere in the presence of points of vulnerability and take a number of measures to address the identified deficiencies; set up a firewall, define filtering rules, as well as to determine the parameters for user authentication.

The integrated system is able to perform the following functions:

- assess the level of network security;
- monitoring of the network;
- traffic management;
- detection of internal and external attacks;
- impaired recovery mode the network.

3. Method of intrusion detection based on control of states

This article focuses on building intrusion detection subsystem based on the method of the states of objects and subjects. At the same subjects and objects of info-communications system (ICS) can be divided as follows:

- local hardware nodes ICS;
- the physical links between nodes ICS;
- channeling equipment ICS;
- node objects - software objects on the nodes ICS, which interact within a single node;
- network objects - objects in the software and external nodes ICS software objects that interact with the ICS node.

Each subject and object in the system is characterized by the state. Status of subject and object - a set of objects ICS having access to them, as well as the characteristics of their current workload.

Subjects and objects can be divided into active and passive sets.

Passive subjects and objects cannot be accessed by other objects, and has an active possibility.

We indicate:

S_{act}– many types of active subjects and objects;

St_a – manypatterns active subjects and objects;

S_{psv}– many types of passive subjects and objects;

St_p – many patterns of passive subjects and objects;

In this case, $St = St_a \cup St_p$;

and$St_a \cap St_p = \emptyset$.

Here the object $st \in St$ of each type of access operation defined. These operations consist of a sequence of elementary operations, so-called primitives.

The transition of any pattern of the subject or object in a dangerous condition, by definition, means the transfer of all ICS in a dangerous condition. This definition of an unsafe condition ICS means that its discovery there is no need to collect information on the status of each object in the ICS at a time, but rather only observe the changes of critical states in terms of protection of information objects.

Assume that the model of information attacks based on the following three main sets such as V - multiple vulnerabilities ICS, O - the set of methods for attacks, and **E** - the set of consequences of attacks. In this element belongs (o, v, e) against **W**, where $o \in O, v \in V, e \in E$, can be interpreted as follows: "The external attack carried out by an attacker using the method of o, which activates the vulnerability v, and leads to consequence "[6, 7].

To view an attack, you can use the graph-scheme which is as follows $G = U \cup R$, where: **U** - the set of vertices, and $R \subset U^2$ - the set of arcs of the graph, and **G** defined by the ratio of $P \in \{R \times W\}$, which each arc of the set **R** assigns to one or more elements of the relationship **W**. Application of **P** allows us to interpret the relationship each arc of **G** as one of the stages of a simulated attack information. At the same time on one arc of $Pr \in R$ can correspond to multiple elements of **W** only, provided that these elements represent an attack that lead to the same consequences. At each vertex of a graph **G** may include multiple arcs only, provided that in respect of each such arc **P** correspond to the elements of **W**, describing the attack, leading to similar consequences. Thus, the vertices of **G** can combine the various stages of the attack, leading to identical results.

An example of a graph **G** is shown in figure 2, which describes an arbitrary information attack, as well as the ratio of **P**, which defines the stages of the attack, modeled by the arcs of G·

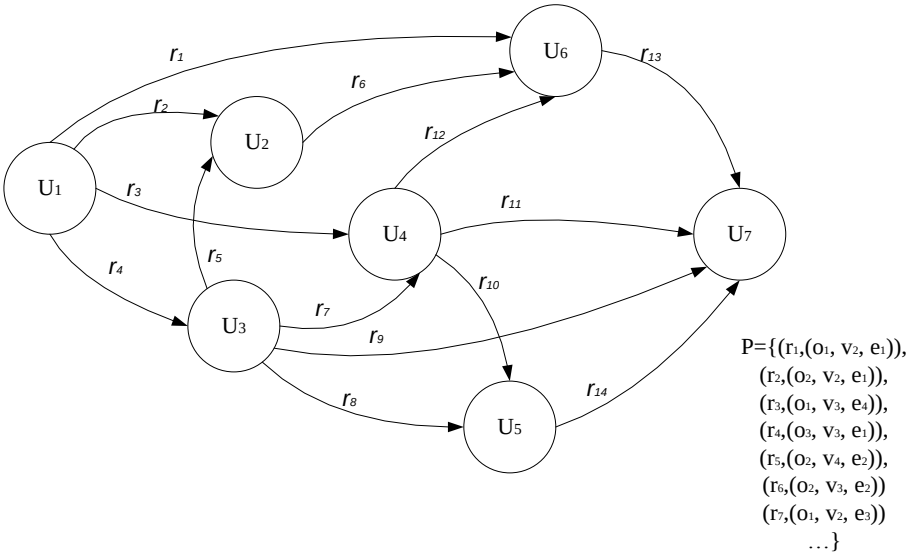


Figure 2. A graph-state transition diagram

The graph G can be represented as a model of information attack, and from this graph can detect network attacks, determining what paths lead to the consequence, for example $e_1 \in E$, which is dangerous for the ICS. Scenarios of attack can be many possible paths in the graph $G - Gp$, where each path $gp \in Gp$ is a sequence of arcs $(r_{p_1}, r_{p_2}, \dots, r_{p_n})$ of the form$r_{p_k} = (u_i, u_j), u_i, u_j \in U$, with the ultimate top of the arc r_{p_k} is also the initial vertex of the arc $r_{p_{k+1}}$. As an initial vertex path may make such vertices $u \in U$of G, indegree equal to zero. Final vertex path can only be a vertex u, outdegree is equal to zero.

Multiple paths for the graph Gp isas follows: $Gp = \{\{(u_1, u_6), (u_6, u_7)\}, \{(u_1, u_2), (u_2, u_6), (u_6, u_7)\},$

$\{(u_1, u_4), (u_4, u_7)\}, \{(u_1, u_3), (u_3, u_5), (u_5, u_7)\} \dots\}.$

Thus, the attack leads to the consequence $e_1 \in E$ can be detected as follows:

-conduct an attack using the $o_1 \in O$, activating the vulnerability $v_2 \in V$ and leads to the consequence $e_1 \in E(gp_1 = (r_1), (r_1, (o_1, v_2, e_1)) \in P)$ or in an unsafe condition $\omega(st) = 1$;

-conduct an attack using the $o_2, o_1, o_2 \in O$, activating vulnerabilities $v_4, v_3, v_2 \in V$ and leads to consequences $e_2, e_4, e_1 \in E$, respectively

$(gp_2 = (r_5, r_3, r_2), (r_5, (o_2, v_4, e_2), (r_3, (o_1, v_3, e_4), (r_2, (o_2, v_2, e_1)) \in P)$ or in an unsafe condition $\omega(st) = 1$;

-conduct an attack using the $o_2, o_3, o_2 \in O$, activating vulnerabilities $v_5, v_3, v_2 \in V$ and leads to consequences $e_3, e_1, e_1 \in E$, respectively $(gp_3 = (r_6, r_4, r_2), (r_6, (o_2, v_3, e_2), (r_4, (o_3, v_3, e_1), (r_2, (o_2, v_2, e_1)) \in P)$ or in an unsafe condition $\omega(st) = 1$.

Attack on the ICS can be represented as a sequence of actions executed by a pattern of the subject or object, or a set of sequences of action patterns of the subjects and objects that [8]:

$$X = x_1, x_2, \dots, x_k, \forall x_i \exists st \in St: x_i \in t_r, \exists\, 1 \le j \le K: \forall\, i \ge j\ S_j \in \delta$$

Thus, an attack is the trajectory of a pattern of subject and object or set of sections of the trajectories of a group of subjects and objects template that displays the ICS from a safe state.

The normal behavior of the object st can be defined as the set of trajectories of all object instances of this type of $type(st)$, which do not derive from the ICS safe condition. Note that none of the trajectory, a part of the normal behavior of the object can not contain any hazardous conditions.

The proposed method of intrusion detection is used many examples and descriptions of trajectories attacks normal behavior for the detection of signatures and anomaly behavior of objects by comparing the observed patterns of behavior of objects with given examples.

In a formal statement of the problem of detection intrusion detection problem is divided into sub-tasks of detection signatures and anomaly detection.

Detection of signatures. Set by:

$H = \{H^i\}, H^i = x_1^i, x_2^i, \dots, x_N^i$ – many examples of attacks;

$T_{sys} = \theta_1, \theta_2, \dots, \theta_N, \dots$ - sequence of observed states the protected system.

Required to find a lot of templates subjects and objects of $O^* = \{o\} \subset St_{act}$ and the corresponding set of trajectories $T^* = \{T_o | o \in O^*$, which can attack from a variety of examples of attacks.

Anomaly detection. Set by:

$B = \{Bh(type) | type \in S_{act} \cup S_{psv}, \forall st \in St: type(st) = Type, \forall\, S \in N_{st}\}$ - many descriptions of the normal behavior of the subjects and objects of ICS, defined in terms of types of subjects and objects.

$T_{sys} = \theta_1, \theta_2, \dots, \theta_N, \dots$ - sequence of observed states the protected system.

Required to find a lot of templates subjects and objects of $O^* = \{o\} \subset St_{act}$ and the corresponding set of trajectories $T^* = \{T_o | o \in O^*$, which do not belong to the description of normal behavior for the respective types of subjects and objects of ICS .

4. Process of detecting anomaly in information and communication system

The process of intrusion detection is reduced to the formation of the observed sequence of states and the ICS finding in this sequence of attacks and the trajectories of deviations from normal behavior:

1. For each protected object is formed by a set of templates $stK_{st} = \{K_{st}^i\}$ (each pattern in this set of models the attack of a class i) and the template N_{st}.

2. If you receive some of the observer of *a*.

-Defines the set of actions to which this symbol belongs.
-Is determined by a subset of templates that can take this character.
-Served as *a* symbol of the input pattern to each of the selected subset.

3. For each pattern is calculated from the selected subset of the predicates of the transition from the current state of the symbol *a* and the predicate of the final state.

4. If the final state of the template is the end, the action specified for the final state (formed by the message of the attack), and this pattern is destroyed if it is not the only one, otherwise it returns to its initial state.

Thus, for the effective functioning of the intrusion detection subsystem and integrated system of information security in general, requires the following specific requirements:
- the presence of an audit trail;
- identification, authentication, security administrator;
- set the terms of the security attributes;
- security of the information sphere;
- self-testing;
- the restoration of health.

5. Conclusion

Using an integrated security system provides the following advantages.
*The centralization and standardization of audit log format.*Detection, firewalls, and security analysis tools operate within a complex integrated system of software that allows you to use a single service logging security audit.

Simplicity of administration. Placement of detection, firewalls, and security analysis tools in a single integrated system allows the use of a common management interface for each system component.

The high operating efficiency. The high level of efficiency of the integrated system is achieved by eliminating one component of the system and fill their other advantages. Thus, the main drawback of detection of unauthorized actions, is the inability to analyze the contents of data packets transmitted at high speed, can be eliminated by using a firewall. Extending the same rules packet filtering firewall can reduce the intensity of the incoming traffic and thus increase the efficiency of detection of unauthorized access to the network. The main drawback of the firewall - inability to protect against sophisticated attacks by conventional methods of filtration - is overcome by the inclusion of an integrated system of detection of unauthorized actions, which has sophisticated algorithms to detect and localize the offending actions both outside and inside the network.

Using a single database of attack signatures. The functioning of these components into a single software package allows them to use the same data bank of unauthorized signature effects, which greatly simplifies the procedure for updating the data bank.

Reference

1. ShirochinV.P., MukhinV.E.,KulikA.V. Design Issues of information security in computer systems and networks. Kiyev; «VEK+». 2000. – 111 p.
2. KovalevskiyV. Cryptographicmethods.// Computer Press, № 5, 1993.
3. GanievS.K., KarimovM.M. «An integrated system of information security in corporate networks»./Theses of reports of the International SCP «Innovation 2001».-Tashkent, 2001.-P.97-98.
4. GanievS.K., KarimovM.M. «Combined method of crypting to protect information systems»./Theses of reports of the RepublicSCP«Technical disciplines and global issues XXI century ».-Tashkent, 2001.-P.12-13.
5. SerdyukV.A. Technology unauthorized actions on the Internet //Application to the Journal «Information technologies». 2001. №5. P. 1- 24.
6. V.A. Serdyuk. Analyze of modern lines of construction of models of information attacks// Information technologies, № 4, 2004.
7. A.A. Klimovskiy. To the analysis of approaches of classification of computer attacks// Materials of the international scientific conference on problems of safety and counteraction to terrorism. –M.: MCNMO, 2006 -480p.
8. N.J. Miloslavskaja, A.I. Thick. Intranet: Detection of intrusions. The manual for high schools. –M.: UNITIDANA, 2001. – 587p.

Information Display Management in Automated Systems

A.F. Verlan[1], Yu.O. Furtat[1], D. Velev[2]

[1] G.E. Pukhov's IMPPE NAS of Ukraine, 15, General Naumov Str., Kiev, 03164, Ukraine
a.f.verlan@gmail.com, saodhar@ukr.net
[2] UNWE, Stud. grad, 1700, Sofia, Bulgaria
dvelev@unwe.acad.bg

Key words: user interface customization, automated system workplace, automated adaptation, cognitive profile.

User interface management – the need and principles

The task of creating information storing forms and presentation adaptation software has emerged as a result of the automated systems quantity, diversity and complexity increase. Increased has not only the amount of information with which the user has to do in the system, but also the number of users with their data processing and perception particularities. Existing methods of enhancing the efficiency of the users' professional efficiency (ergonomic workspace organization, purely organizational methods, etc.) no longer give satisfactory results.

Consequently, there was a need to supply the user with means to modify the workplace interface depending on the task features and objectives.

The main criterion for optimal customizable workplace interfaces functioning should be convenient user's reception of information from the system, speeding up decision-making based on this information, and reducing errors. Meanwhile no important information should be lost or delayed – in case of critical situations in powerful objects management systems it could lead to a disaster.

User interfaces adaptation software for complex systems should take into account the user's particularities, as well as his wishes regarding system information on-screen presentation form and specific systems features and their goals.

Typical user interface of any complex system consists of a common elements combination – text information display fields, graphs, scales, tables to summarize the values of various system components. Currently the vast majority of interfaces are of the same type – a combination of graphical and textual information representation means. This form is the most universal, but cannot take into account the individual characteristics of all system users.

Since most such schemes are the result of the initial system information processing, they usually exist in form of text or digital data blocks. Thus, tuning the interface does not mess with the information, it simply recombines it to represent in the most convenient form for the user's perception.

User interface customization principles should be similar to the ones used when creating a form in such environments as Delphi or C++ Builder. Available to users are different output elements grouped into separate panels, for example, by the information presentation form. The user selects an item and places it on the interface field, also

connecting it to the data source. However, at least one output component for general information must be present on the form – to prevent the loss of potentially critical data.

Thereby, configurable user interface could be realized in form of several separate auxiliary panels containing output elements with the possibility to consult the complete system information, presented in one of the most popular types of user interfaces (mnemonic diagram, data blocks tree, etc.), at any time.

Automated information presentation form management

The interface adaptation system should also be able to choose the most user-friendly form of information presentation. For this the functions of the automatic part of the system should be expanded.

First, the system must retrieve the user's information perception characteristics from his cognitive profile. This profile can be provided from outside the interface adaptation system (e.g., through professional psychological testing) and by the system itself – by conducting tests at the workplace and/or monitoring the user's behavior and performance during his professional activity.

Second, the interface adaptation system must contain data about the specifics of the process with which it is connected.

Third, there must be means to analyze files containing information from the automated system to choose the most appropriate set of output elements.

Systems specialties' influence on interface adaptation process

As it has been already indicated above, for the information flows characteristics adaptation to the peculiarities of users perception the automated systems' intellectuality is used.

However, because of automation level increase additional user interaction problems have emerged. The problem may lie in excessive trust in automation - then the user loses vigilance and cannot react in time to the critical situation thus reducing overall system reliability. If the user, on the contrary, does not trust the automatic components (main reason – because their behavior strategy and goals are hidden from him), he spends time and resources to double-check system actions or could even turn off automatic equipment (such situations are frequent in aviation – [1]). This reduces the system effectiveness and then, through increased user fatigue, its reliability is suffering.

To solve this problem, user could be given access to the automated system to control the intellectual components' behavior and receive information about the selected optimization strategy and its current steps. Possibility of intellectual components control should depend on the system characteristics and the user's training level. For example, in automated control systems (ACS) user is a highly trained specialist, which is responsible for his decisions. He may be provided complete control over the means of adaptation and automation. In education system user is most often a student. The system has to effectively lead him to his goal, user intervention is unprofessional and will result in reduced system efficiency. In this case the system should only inform user about the system behavior strategy – understanding it could be an additional motivating factor in learning.

User information important for user interfaces adaptation process automation

Formal characteristics of the message include: content, form, rhythm and pace of information (the spacing between individual reports, regularity of income, etc.) [2].

Results of experimental studies indicate that among five main information processing stages representation of the initial situation is a special case [3].

Thus, the presentation of information can be viewed as a way of to manage the system user's cognitive activity because it is an essential factor that provides the perception, understanding and mastering the studied material. It is known that while mastering the material in the learning process the person selectively refers to its contents, type and form [4].

In addition to the preferred system information on the workstation presentation form of information, each user has a number of characteristics that affect user-directed information perception, comprehension and mastering.

These characteristics are divided into cognitive, psycho-physiological, and intellectual.

Cognitive characteristics mainly determine the speed at which the user orients himself in the new information and switches between different types of activities. These include:

- field independence-field dependence;
- narrow-wide range of equivalence
- impulsivity-reflexivity;
- intellectual lability-rigidity.

Apparently, the cognitive characteristics are presented in pairs of values, which are extreme manifestations of the corresponding characteristics. Therefore, for a specific user value of each performance is between 0 and 1 depending on the inclination to one end of the scale.

To assess psychophysiological state characteristics are often used such as the user's attitude to risk, the quality of memory and attention. At the same time, the reaction to the test impact and response bias should allow to assess concentration and coordination of the user, as these depend on the personal particularities.

Intellectual characteristics are the levels of IQ and EQ (intelligence and emotional intelligence quotient, respectively). Although the coefficient associated with emotional intelligence depends on IQ, there is no direct relationship, namely the value of EQ determines predisposition to creativity when working with information. This characteristic is very important to users of design or educational systems.

Lexical analysis of the input information to manage the presentation form

As mentioned above, the initial information from the system has an array of numerical data or structured text. To present this information in a user-friendly form it should be divided into the content units, and then each such unit should be transformed to the appropriate form and displayed on the corresponding widget.

To perform breaking into the content segments user interface management software has to analyze file with information from the system (working directly with the information

flow from the system in real time is undesirable because of potential problems with synchronization with the real-time adapted interface) based on key words that mark the beginning and end of certain segments of the file, i.e. conduct its lexical analysis.

Nowadays, in most systems use the established format with a specific set of keywords. A set of keywords depends on the type of file, but users may be allowed to manage a set of the analyzer's tokens. Also, the tools are needed to manually format the text when the original formatting is incomplete or missing.

The best option would be to implement the lexical analyzer as a separate program, and its results to be entered into the database (permanent or temporary), where data can already use the software management interface. This will further insure against the synchronization problems.

Working with semantically-sensitive information

Often, for a better perception of information it is expedient to give to the user not only the required at the moment information, but also support or help information, associated with it in meaning. In the general case, the information is stored in separate blocks of data. To correctly display semantically related or dependent information it is necessary to establish the relationship between such blocks and to name them to provide selective access.

Thus, from the concept of the data, we turn to the concept of knowledge (which, in the simplest case, is named blocks of data with applied relation). And it is with the knowledge base user interfaces management software must work for efficient information presentation form manipulation.

Working with the knowledge, organized into the knowledge base, requires using intelligent software technologies. The user-system interaction is thus regarded as a dialog interaction in which both parties are agents of dialog and act on corresponding scenarios. Dialog interaction scenario depends on the subject area specifics and user skill level. For example, in the case of highly trained user part of background information may be hidden (to streamline interface) or displayed only by request (supported text document - [5])

Algorithmic description of one of the possible variants of interfaces management system realization

Now that the basic principles of user interface management are formed by the scheme and the principle of such a system can be described.

Figure 1 shows the path that information passes in the system with a customizable user interface – from gathering information from the system and to the output on the screen user workstation screen.

Final presentation of data user gets on his workstation. Information about the objectives and the current system adaptation process stage should be displayed in a separate Help window along with information from the system. The research [1] showed that user awareness of the user interface adaptation system actions increases his confidence in the system, and it is more likely for him to positively perceive manipulation with user interface and presentation form.

Lexical analysis is an initial stage of system information processing. It has several sub-stages, and lexical analysis itself (i.e., finding known tokens in the system data file) is preceded by equally important phases of the file structuration – fig. 2.

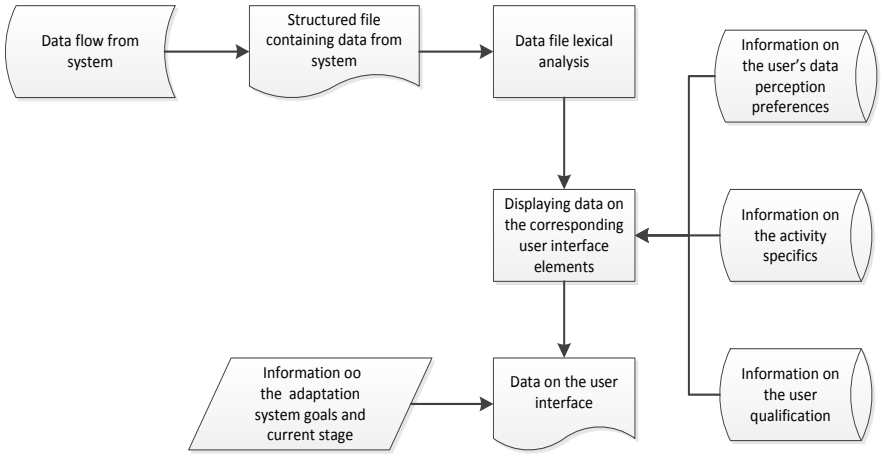


Fig. 1. The processing of information in the system with a customizable user interface

It should be noted that breaking the file with information from the system into the content elements must take into account not only the structure of the file (tokens), but also the current state of user interface. In particular, you should not break the file into more content elements than there are elements in the output interface. If the user does not have the necessary element to represent particular content elements form the output, you need to take additional steps - inform the user (in case of user interface manual configuration) or to add an appropriate user interface element (in the case of automatic adaptation).

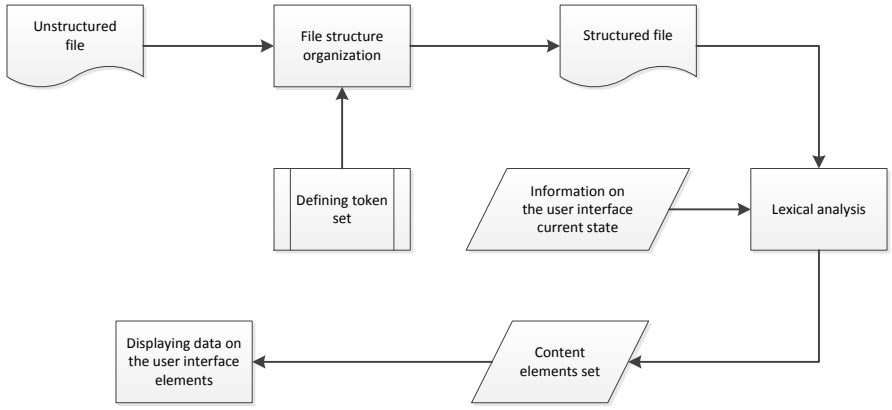


Fig. 2. Data file structuring and lexical analysis

The Fig. 1 shows that the process of information on-screen output should take into account the user's peculiarities of information perception, his skill level and professional action specifics. It is generally taken by interfaces management system interfaces from user profiles prepared earlier by external experts (psychologists, members of staff, engineers, etc.). But these features of the user's information perception can also be received from the system, allowing to adjust the initial cognitive profile (in this case still unresolved ethical issues associated with automatic evaluation of training and fitness do not occur). Fig. 3 shows the automatic testing system scheme.

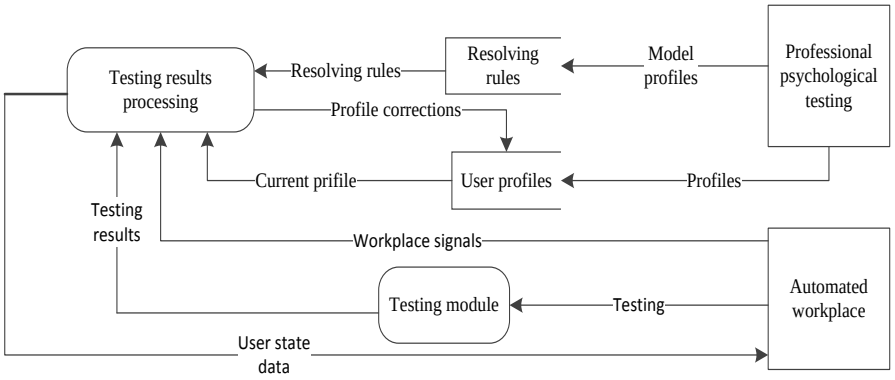


Fig. 3. Scheme of the automatic user characteristics diagnostic system

As can be seen from the scheme, user testing can be performed explicitly (user passes tests) as well as in the background when the workstation tracks metrics such as speed of action or reaction to situation change. In both cases, information about the characteristics of the user is treated by separate testing module and supplements or changes the initial cognitive profile for further use in user interface management system. This way – by altering the user interface and observing the change in his performance (which is expressed in increasing the reaction rate and reducing the number of errors when working with information), it is possible to optimize interface for a particular user with a finite number of steps of an iterative process

Conclusions

Based on the provided operation principles description and an example implementation working user interface adaptation software may be created that will take into account all peculiarities of this problem. The use of modern achievements of the theory of artificial intelligence and intellectual software will make this system flexible and truly adaptable. Using cognitive approach (aiming at increasing the efficiency of the process of information perception and processing by the user) will reduce the load on the user and increase efficiency (and reliability) of an automated system as a whole, because user is its important functional part.

References

1. Christopher D. Wickens, "Imperfect and Unreliable Automation and Its Implications For Attention Allocation, Information Access and Situation Awareness", Technical Report ARL-00-10/NASA-00-2, Aviation Research Lab Institute of Aviation in University of Illinois, 2000. – 28 p.
2. Турзин П.С., Пономаренко В.А., Рысакова-Ромашкан С.Л. Уровни понимания информации и структура коммуникативного акта. – Психологический журнал, 1992, Т. 13, № 1. – С. 30–39.
3. Strenberg R.J. Inside intelligence – Amer. Scientist, 1986, V. 74, № 2. – P. 137–143.
4. Якиманская И.С. Принципы построения образовательных программ и личностное развитие учащихся – Вопросы психологии, 1999, Т. 13, № 3. – С. 39–47.
5. Chmyr Igor, "Dialogue of Partners as a Method For Non-Formal Problem Solving" in Maddy D. Brouwer-Janse and Thomas L. Harrington (Eds.) collection: Human-Machine Communication for Educational Systems Design, NATO ASI Series F129 Springer-Verlag, Berlin, 1994.

System of Monitoring Security Information and Communication Systems

Madjit Karimov, Miraziz Sagatov, Abdukhalil Ganiev and Dilmurod Abdullaev

Tashkent university of information technology, str. A.Temur, 108, 100202, Tashkent, Uzbekistan, dr.mmkarimov@rambler.ru, k.akhmatovich@gmail.com

Abstract. This article analyzes the means of monitoring security of information and communication systems. A model and architecture of the monitoring system security includes four modules for the control of objects and subjects. Proposed architecture, a model and an algorithm for constructing systems to monitor the user experience of information and communication systems based on collecting and analyzing data journaling. The main objectives of the analysis examined the problem of determining the statistical parameters and characteristics of users and the construction of the generalized profile, and the problem of early detection of intruders and internal analysis of the traces that have already happened attacks.

Key words: Information communication system, monitoring system, information security, computer networks, model, secure monitoring system, algorithm, earlier detection of attacks, monitoring of works.

1. Introduction

For organizations, computer network (CN), which comprise, more than a dozen computers with different operating systems in the first place stands the task of managing a diverse array of defense mechanisms in such heterogeneous enterprise networks. The complexity of network infrastructure, a variety of data and applications lead to the realization that information security beyond the security administrator's attention may remain many threats. That is necessary to conduct ongoing monitoring of security of info communication systems (ICS) [1].

Over the past decade in solving the problem of secure access to resources, information and communication systems can note significant changes. More recently, the security of information and communication systems can be reliably ensured by such traditional protective mechanisms, such as identification and authentication, access control, encryption, etc. [2]. But with the advent and development of open computer networks the situation has changed dramatically. Connecting corporate networks to the Internet to build distributed networks, the emergence of a huge number of viruses actively promoted the implementation of technical means to protect the perimeter of the ICS. Recent studies show that today 90% of companies use firewalls and antivirus programs, and 40% - Intrusion Detection System (IDS). The large number of different security tools generates a large number of logs from different systems, each of which produces the event in its own language and, moreover, gives a lot of unnecessary information, and sometimes works at the event is not quite adequate [9, 3]. In this regard, one of the most urgent tasks is to build a unified, decentralized price-monitoring system of information security events.

According to the research priorities for monitoring of virtually all companies are:
- define the parameters of the user experience;
- "Early detection" internal intrusion;
- identify signs of internal intrusions.

2. The existing approaches

Purpose of monitoring systems - is to detect anomalies in the ICS and promptly respond to them. Detect suspicious activity is subject to the system components - from the users (both internal and external) to the software and hardware. The existing monitoring system is used to detect attacks and suspicious activity to the ICS or the network or system approach [4, 5].

When monitoring the network layer is used as a data source for analysis of raw network packets. As a rule, in monitoring the network layer uses the network adapters that operates in the «listening», and analyze traffic in real time as it passes through the network segment. Attack recognition module uses a method of detecting attacks, and as soon as the attack is detected, the module notifies the responding attack or triggers an alarm.

Monitoring system-level controls system, events and logs security events. When any of these files is changed, then there is a comparison of new entries to the signatures of attacks to see if there's a match. If a match is found, the system administrator sends an alarm or activates other specified response mechanisms [7].

Each level of monitoring a CS is different and covers a specific area of action, the results of the comparison of options for monitoring at different levels is presented in Table 1.Table 1.

Table.1

№	Thelevelofmonitoring	Type	Advantages	Disadvantage
1	The level of system	1. System 2.Developments 3. Journal of security	1. Confirms the success or failure of the attack; 2. Supervises the activities of a particular node; 3. Detects attacks that network-layer system are missing; 4. Well suited for networks with encrypted and switched; 5. Do not require additional hardware, 6. Low cost of operation.	Depends on the OS and does not detect the anomalies arising at the network level

2	The level of network	1. The analysis of network packets 2.Traffic analysis	1. Low cost of operation; 2. Detects attacks or anomalies in the behavior that occur at the network level; 3. Detects and responds in real time; 4. Detects unsuccessful attacks or suspicious intent; 5. Do not depend on the OS.	Cannot detect system-level attacks

3. Model and algorithm of monitoringsecurity

In this article we propose a centralized data collection and analysis, as this approach has the following advantages:

- Allows for the collection of evidence to investigate the violations occurred already;
- Improving the quality of analysis by the correlation of data from various sources;
- Improving the visibility analysis by displaying the information at once about the whole network or subset;
- Less congestion and client nodes, as a consequence, the ability to create "invisible" to the user agent acquisition.

3.1. Model system for monitoring the safety can be described as anomaly detection based on the profile. Profile is based on the results obtained during the monitoring of information and communications network.

Formally, the profile described by the tupleP_{ICS}:

$$P_{ICS} = \{W_i, S, S_S, N_E\} \quad (1)$$

where:

W_i – number of information units, which are considered as the data obtained in the analysis of security logs user workstations;

S – number of information units, which are considered as the data obtained in the analysis of log-file servers and service availability of the CP (DBSC, Web-server, proxy-server, mail server);

S_S – set of information items about the operation of security services;

N_E – describing a set of network equipment, i.e. equipment responsible for the operation of corporate networks, such as routers, switches, hubs, communication lines, etc..;

At the same time identify two modes of profiling:

1. Preparation of the reference profile of ICS (P_{ICSe}), the profile is formed by the test monitor and characterize the state of the CP, which was originally considered to be safe.
2. Preparation of the current profile of ICS (P_{ICSt}), the profile is created each time the monitoring of the CP.

The next step is to compare two profilesP_{ICSe}, and P_{ICSt} which resulted in the conclusion of the anomalies in the system, which may indicate a security breach in the ICS.

3.2. Algorithm of monitoring security.

At the core module for monitoring the security of ICS is the preparation of the reference profile and the drafting of the current profile, which is then compared with the reference.

A block diagram of a monitoring system algorithm is shown in Pic1.

Sequence of the algorithm:

1. The beginning of the algorithm.
2. Check whether there is a reference profile.
3. The profile is loaded.
4. If there is no profile reception of parameters of monitoring is carried out.
5. On the basis of the received parameters of monitoring it is formed etalon profileP_{ICSe}.
6. Data processing process begins.
7. Reception of parameters of monitoring is carried out.
8. On the basis of the received parameters of monitoring current profileP_{ICSt} is formed.
9. The current profile is compared to a reference profile.
10. If the comparison is, the decision of the monitoring system is the absence of anomalies and irregularities.
11. If comparison does not match, the security administrator is given the message that there is a security breach.
12. The end of the algorithm.

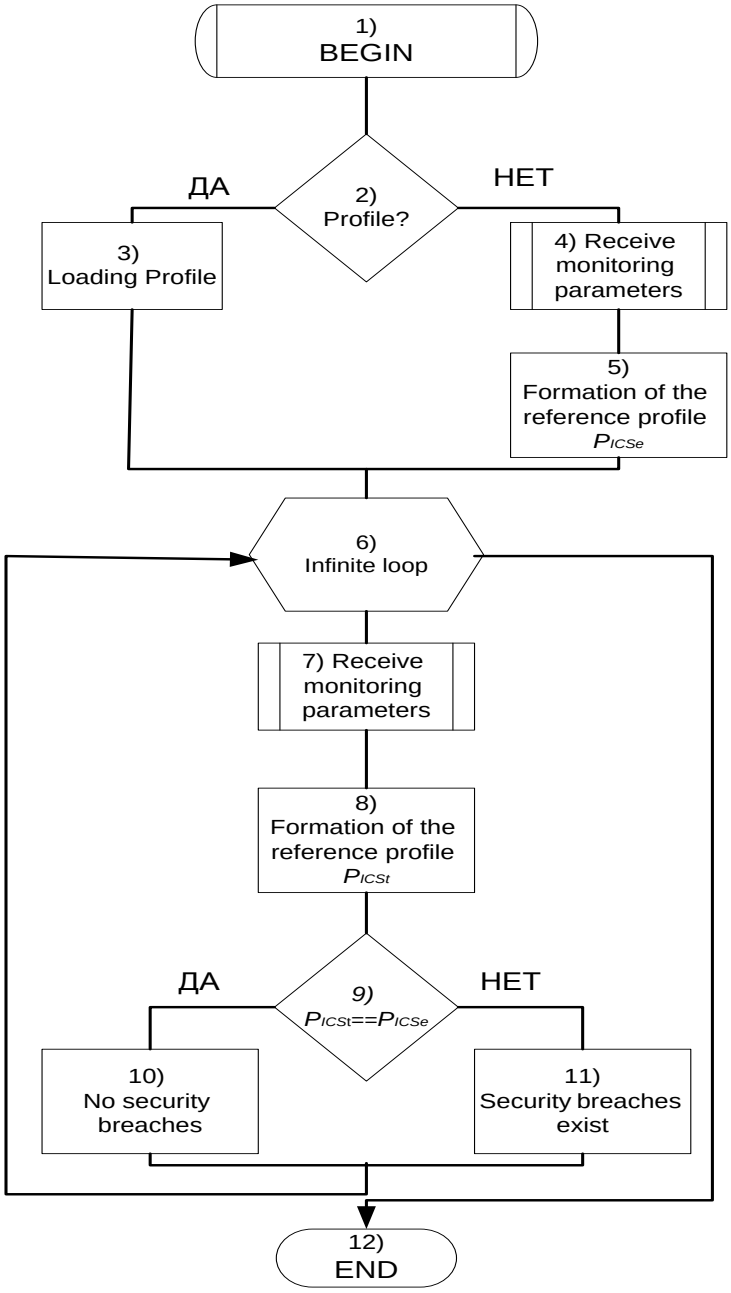


Fig.1. Block diagram of the algorithm monitoring system

4.The system architecture of security monitoring

Architecture of the monitoring system has several modules for data collection and analysis of the states of objects and subjects of the information and communications system (pic.2).

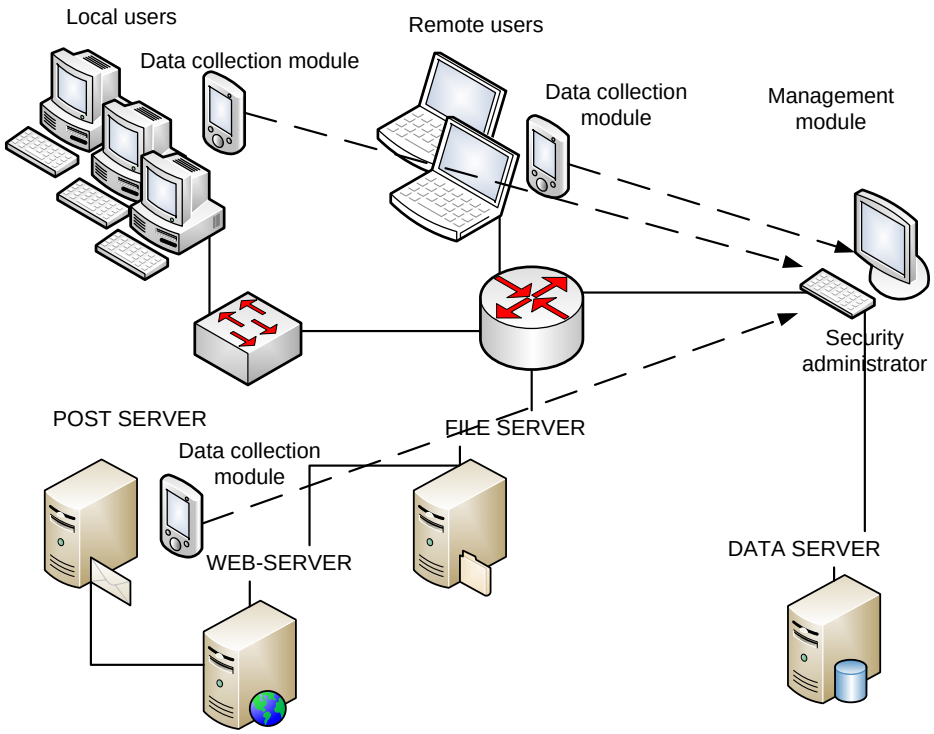


Fig.2. The system architecture of security monitoring.

The architecture comprises modules of data collection, data storage module, control module collects and workplace analyst.

The developed architecture allows modules to operate in a collection of stand-alone mode; the modules may store data in local storage, and transfer them to a storage module later. The architecture allows flexibility to configure the collection of initial data, namely a set of magazines, a set of events, filtering records events using regular expressions. The architecture allows you to adjust the policy data, as well as view the status of system components and automatically detect faults in the modules.

One of the most important requirements of a system for monitoring the user experience is its performance. In today's ICS, numbering thousands of computers, data flow logs can be overwhelming, especially if we take into account the availability of logs, which

record is at a rate of ten events per second. Centralized analyses of such data are computationally difficult and sometimes impossible task [8].

The first obvious step to reduce the flow of data is to filter data during the data collection, however, as experience shows, only the filter is usually not enough.

You can use the approach of load distribution analysis, which consists in the formation of the modules collection of facts describing the activity of users. Evidence of activity generated by the specified rules based on the analyst's collected.

This approach allows the analysis of the spread between the modules to collect and workplace analytics, to unify the analysis of the data, as already analyzed the facts, not arbitrary log entries, and as will be shown below, usually increases the visibility and quality of analysis. Also, this approach reduces the flow of data from the agents and the load on the data storage unit.

The component of the system implements the following basic functionality:

1) The gathering module;
2) The storage module;
3) The module management of gathering;
4) A workplace of an analyst of security.

4.1. Data collection module - reads data from the logs OS carries out pre-processing of collected data (the formation of evidence of activity) and in accordance with planning policy transmits data to the server consolidation [6].

We can formulate the following requirements for data collection module:

- «imperceptible» for users of the system's reading of arbitrary text, and the system logs;
- collection of additional activity, not the standard journaling log files;
- the implementation of the mechanism of filtering events by event type, and the values of event arguments;
- creating events on the basis of facts read by the activity;
- setting policy transmission across the network;
- to protect themselves and collected, but not yet sent data from damage / theft / spoofing.

4.2. Storage module – collects data from the module collection and puts data into a common repository. Getting data from the module, storage module converts records into an internal format and stores them in a store. As part of the proposed unified view of the analyzed events (evidence of activity) in a system that allows you to consistently provide evidence of activity for a unified storage and analysis, which does not depend on the type of the original log.

The proposed data warehouse can be effectively collected evidence of activity with the use of dictionaries of real values of the arguments of events. Storage is built on a file system that provides greater productivity as the most important parameter of the storage unit and storage in general is just performance. The load on the repository consists of load to add the newly received data journaling, and the load on the extraction of data for analysis. The nature of these actions is different. Data extraction typically involves rapid receipt of all entries collected for a certain time period, and this problem is solved by placing entries in

sorted order by date. It is more difficult with the addition of new data obtained as a server consolidation to combine the data from a large number of modules.

Performance problems can be solved by splitting the files store information in three parts:

- the main parameters that are present in almost all records of all logs: Event type, generation time, etc.;
- optional parameters that describe additional information about the event;
- directory to store the actual values of both core and auxiliary parameters. To access the values of the reference books used by the hashing mechanism.

4.3. The control module collection - allows you to add, delete, configure modules collect, visualize the status of the collection of modules and module storage (database).

4.4. Workplace Security Analyst - allows you to request data from the database, filter them, fill out sections of data, to perform various types of analysis.

5. Conclusion

In this article the technology of specialized analytical systems for monitoring corporate users. We propose a unique treatment approach journaling of events, based on the formation of the facts describing the activity of users. The proposed approach can significantly reduce the flow of data for analysis, which, in turn, reduces the load on the network and data analysis, no significant decrease, and sometimes with the improvement, quality and clarity of analysis. The technology allows unifying the collection, consolidation of data from various sources, to produce a unified statistical and predictive analysis to determine the parameters of the user experience and search for anomalies in the work. Centralized analysis of isolated facts with the help of data mining algorithms can detect new and modified the known types of intrusions, and locate anomalies in the work of users. Scalability is achieved through the creation of new agents for the required platforms. The proposed solution was tested in a multi-agent system consolidation and analysis of logs in a number of government and commercial organizations.

References

1. Computer Crime and Security Survey [Electronical resource] : review of computer violations and methods of combating them / Computer Security Institute, 7 April 2002y. – Access mode: http://www.gocsi.com/press/ 20020407.jhtml, freely.
2. Cristina Abadyz, Jed Taylory, CigdemSenguly, William Yurcik. Log Correlation for Intrusion Detection: A Proof of Concept. Department of Computer Science, University of Illinois at Urbana-Champaign, 2003. – P. 3–6.
3. Kathleen A. Jackson. Intrusion detection system (ids) product survey. Distributed Knowledge Systems Team Computer Research and Applications Group Computing, Information, and Communications Division Los Alamos National Laboratory Los Alamos, New Mexico USA, 1999. – P. 6–22.
4. National Survey on Managing the Insider Threat [Electronical resource] :review of internal control invasions / Journal.ComputerWorld, 12 September 2006y. Accessmode: http://www.computerworld.com/action/article.do?command= =viewArticleBasic&articleId= 9003211, freely.

5. Theuns Verwoerd, Ray Hunt. Intrusion Detection Techniques and Approaches. Department of Computer Science University of Canterbury, New Zealand, 2002. – P. 2–14.
6. SQL Server 2005 Books Online. Microsoft Association Algorithm [Electronical resource]: SQL Server 2005 Books Online. An algorithm for constructing association rules from Microsoft. – Access mode: http://msdn2.microsoft.com/en-us/library/ms174916.aspx, freely.
7. LukatskyA.Detection of attacks /– SPB.: BXV–Petersburg, 2001y.–624p.
8. S.V.Troshin. Monitoring of corporate users. Questions of modern science and practice.№ 2, 2009y. TGTU, Tambov.
9. M.Romanov. Security of corporate networks: monitoring, analysis, management. Journal «Storagenews». №1, 2007y. Moscow.

Constructing the definition of the concept of "information"

Oleg Malyshev

The Institute of Mathematical Machines and Systems Problems of National Academy of Sciences of Ukraine,
Academician V. M. Glushkova Avenue, 42, 03680 Kiev, Ukraine
oleg_malyshev@ukr.net

Abstract. The actual situation with the concept (term) of "information" can be characterized by following assertion: nowadays many people believe that they exist in so called "information society", but a single common understanding of the concept of "information" is absent. It seems that this is due to the lack of theoretically grounded definition of such concept. The process of construction of adequate definition (interpretation) of the concept of "information" should take into account the next few considerations: 1. Knowledge is the primary thing, information is the secondary one. 2. There exists a fundamental difference between an object (real or abstract) and its description, which sometimes needs to be expressed. 3. Any object, intuitively considered as "informative object", exists in form of real object in distinguishable significant states, defined within the technologies of its creating/using (in defined form). 4. "Informative object" can not be created/used without direct or indirect participation of consciousness, for which the information is or "product" (output), or "food" (input). As result, the concept/term "information" could obtain the following definition - this is the description of something/someone, presented consciously in the defined form.

Keywords: form, knowledge, thinking process, description, consciousness

1 Introduction

In the modern world the term "information" has acquired obviously the broadest distribution. Rapidly developing information technologies invade the various aspects of modern life. The importance of 'information' is highlighted at the international level by the introduction of even a special World Information Society Day (March 27, 2006 UN General Assembly adopted a resolution which proclaimed 17 May as World Information Society Day.). But here is what is strange - a single common understanding of the concept of "information" is absent - see, for example, article 'Информация' in Wikipedia (rus.).

The current status of the concept of "information" is adequately and comprehensively reflected in the fundamental paper [1], where, in conclusion, the authors honestly admit that they can not specify a common understanding of the concept of "information": "There are many concepts of information, and they are embedded in more or less explicit theoretical structures. In studying information it is easy to get lost, i.e. to loose one's orientation". However, the need for it exists: "Building a scientific network as a self-reflective activity presupposes the task of clarifying some common concepts. One of them is the concept of information".

We believe that the concept of information should be clearly defined theoretically, regardless of the areas of human activity and applications, and justified practically. We believe that theoretical approuch to construction of needed definition (interpretation) should take into account the next few considerations:

1. Te theme of "information" can not be fully considered, if not addressed the theme of "knowledge". Hardly anyone would argue that the knowledge and information - are one and the same. But the vision of the relation between them can be controversial. Of course, "knowledge" and "information" are just words. But these are the words that we use every day - why not assign to each of them a certain meaning? Knowledge is the primary thing, information is the secondary one.

2. There exists a fundamental difference between an object (real or abstract) and its description, which sometimes needs to be expressed.

3. Any object, intuitively considered as “informative object”, exists in form of real object in distinguishable significant states, defined within the technologies of its creating/using (in defined form).

4. “Informative object” can not be created/used without direct or indirect participation of consciousness, for which the information is either “product” (output), or “food” (input).

Based on these and some other considerations, we aim to give the theoretical definition of concept of "information", showing the process of its construction.

2 A little physics

Physical world that surrounds us, and part of which we ourselves are - it is something that we with some reason, based on experience or ideally, have or are inclined to divide into separate objects. Having as basic the concepts of "space", "time", "state" and not without reason considering that to them some measures may be applied we could extend on some part of the World the notion of “form”, i.e., the real object, which is a some space-time volume, the state of which is continuously changing. Changes that affect the real objects are the results of interactions between them. Modern physics distinguishes now four types of fundamental interactions: strong, weak, gravitational, electromagnetic. Modern physics does not distinguish between them the so-called "information interactions".

For attempts to insert into the physics the concept of information we can make in first approximation such a simple barrier: it is well known that the computer claims to be the keeper and converter of "information", but if we will look at it objectively - no "information" out there, but only the possibility of using a large number of standardized states of "memory" (RAM, external), and the possibility of very rapid transformation of these states on the basis of fundamental interactions in accordance with the existing algorithm, which also is represented as certain state of "memory".

3 Some philosophy

Consideration of the concept of "information" leads inevitably to the door to philosophy. And if we set the task to consider it together with the concept of "knowledge", then it is just impossible not to define our philosophical position.

By some, and we must say, not clear reasons it is so happened that the concepts of "information" and "knowledge" were embedded in familiar and intuitively shared by many

structure “data – information – knowledge - wisdom” (DIKW) (No wonder that [2] in relation to it uses the adjective «ubiquitous».), in which knowledge is represented as a special kind of information. Unfortunately, our own searches for the sources where a sensible criterion for separation of knowledge of the rest of the information has been given as well as the independent attempts to formulate such criteria [3] were failed. This is not surprising: if we do not really understand what “information” is, then to any example of information that our intuition accepts as an example of information, we can always find an angle of view on it, which presents it as some knowledge.

Someone who does not want to endlessly rotate in this vicious circle must to invert for himself the pyramid DIKW [5] and accept the primacy of knowledge, which is considered as the potential ability, inherent to any object, to interact in a certain way with the environment, and to manifest itself in such interaction [4]. Thus, elevating the concept of "knowledge" to the level of primary, so-called "philosophical" categories, we get a chance to separate “information” from “knowledge” in some intelligible way.

4 Focus - Man

The terms "knowledge" and "information" were invented by man. So let's start our discussion with him - human being. From a “morphological” point of view the such creature is a conglomerate of interpenetrating components (Speaking of "components" we do not divide the original essence on the independent non-overlapping parts. In particular, some its knowledge is allocated to all of it.): organs of reproduction, sensory organs that allow individuals to obtain some knowledge about the environment and about themselves, organs of action, allowing individuals to carry out the environmental impacts and the impacts to themselves, bodies of memory, providing the ability to maintain long-term specific states, organs of government (central and peripheral).

Knowledge of such entity, in particular, includes: reproduction technology (The term "technology" used in its primordial sense – this is a knowledge about how and what to do to achieve the desired result.), technologies of identification of real objects and their states through the sensory organs, technologies of maintaining of viability through the implementation of the exchange processes with the environment.

5 The thinking process as an object of cognition and mapping

One of the important factors determining the variability of the knowledge embodied in a person, it is the thinking process flowing in it or associated with it (We carefully avoid too categorical localization of this process, as we know and the saying of V.V. Nalimov "Man in a deeper sense thinks with his body" and, especially, the statement, attributed to Heraclitus, "The power of thought is out of the body."), which, as we believe, can be considered as a real object. Whatever model do we construct about the organization of this process, any of them should allow the presence in it of all sorts of objects that are visible by inner eye and identifiable by consciousness of the person. In particular, the participants of this process are images of real objects with which the individual was faced earlier, impressions of minutely perceived real objects, internal fantasies about objects, actual or potential reality of which may be a separate issue. But this range of internal representations could not be limited, because in addition to the real or pseudo-real objects in thinking

process the representations of abstract objects, the statements of own feelings, the registration of certain events, etc., are present. The monitoring of the process of thought, carried out by consciousness, allows recognize and identify these and other participants.

A special place in the thinking process occupies the generation of object descriptions, which the person for some reason may have/want to make real. The available internal representation may also be the object of internal testing to determine whether the object has the necessary qualities. Of course, the level of quality of the generated image as well as of its testing depends on intelligence and experience of person.

6 Human interactions

Of course, we can not cover by our review every type of human interactions (man with himself, man – other man, man – environment, etc.) and we do not pose such a problem. We consider only the following typical interactions:

6.1 Perception

Existence of the organs and technologies of perception gives to the person the ability to map in its own processes of thought certain aspects of the certain fragments of the World Picture available to it (We emphasize that this is a two-stage system of constraints: 1) a person "sees" only a very small part of the World, and 2) even what he "sees", he sees very limited.).

Aware of its own capabilities and limitations, feeling the needs to extend them and in some way developing their design skills and capabilities of impact, a person invents and uses the new technologies of perception. In this case:

- mapped aspect of the “picture” fragment is included in the scope available for human perception, but goes beyond the available range. For example, the “picture” is too small or too far and some augmentation of the image is needed – so, we use microscope, telescope;
- mapped aspect of the “picture” fragment is not included in the scope available for human perception. For example, the characteristics of the electrical processes, internal product defects – so we use the ECGs, encephalograms, projections of the body in the X-rays, etc.

The result of applying of the technology always comes to man in the form determined by the technology. With this there are two options:

- the form allows a direct perception by the senses of person. For example, inventing a microscope (telescope), the person, putting it between itself and the object of consideration, just changes the scale of pattern, which he/she is still seeing with his/her own eyes;
- form does not allow a direct perception by the senses of man. These forms are usually used with other technologies, and there it is always possible if necessary to acquaint people with its content.
- Another practical aspect that requires special technologies - fixation of static or dynamic images (processes) with the possibility of their re-displaying (for example, photo, video).

6.2 Impact

Existence of the organs and technologies of impact gives to the person the ability to influence on certain available aspects of itself and on some fragments of his environment.

Aware of its own capabilities and limitations, feeling the needs to extend them, a person invents and uses the new technologies of impact. The result of impact, as a rule, can be interpreted as a form. The process of impact, to which some impulse (external or internal) gives the start, can be considered as some integrity, which is not subject to decomposition, or it can be considered as two-stage process, where at first stage the image of the target form arises. In turn, here are two possibilities:

- this image is only a participant in the thinking process and doesn't go beyond it;
- this image is embodied in some form using a particular technology.

6.3 Communication

One of the most important aspects of human life is the communication between human beings or, sometimes, between human and other living beings. In principle, the communication can be considered too large - for example, we could include to this the interaction, necessary for procreation, or a fight, as the component of deploying process of conflict. But we do not do this, limiting the interaction by the processes of thinking.

If the processes of thinking, in turn, are thought of as real objects, in order to interact, they must somehow overlap in space/time. Perhaps, the Nature provides a mechanism for allowing the direct interaction of the thinking processes (so-called "telepathy"). But, since we do not have the reliable data on this, then we have only to organize such interaction using the objects-mediators, serving carriers of the thought process fragments maps.

Of course, such role of an object-mediator can be played by the individual itself. If we consider the living being as an object, with which the process of thinking is associated, then, as a rule, this process might have some manifestations caused by the presence of by one or another way realized dynamic relationships between the participants of the thinking process and the impact organs of thinking being. These manifestations can occur unconsciously or consciously. In the case of awareness they can be for various reasons false. For example the "mechanism of lying" may be like this: I feel the emotion "X" (awareness of the fact) → I do not have to show emotion "X" → to hide the presence of emotion "X", I try to portray «Y». Again, there is an intermediate description of how the next step should be implemented.

But we know that for the mapping knowledge contained in the process of thinking, man has learned to use not only themselves, but also the objects external to him. With his initial technological capabilities (gestures, grimaces, postures, sounds), he was able to implement the "promotion" and create a lot of new technologies that map the fragments of thought process in some defined form.

From a motivational point of view, there two options are possible:

1) he/she wants to fix some of his/her knowledge in order to save it from possible loss (everything flows, everything changes - in particular, is destroyed, forgotten);

2) he/she wants/must/have to share some of his/her knowledge with the external living world.

In both cases, a human being H must somehow identify the part of his knowledge. Let X - the identifier of the knowledge of H, which in turn is a part of H. X can be represented as an operator which when applied to H, sets focus on a certain part of it - X (H).

H can neither to separate from himself/herself that part, nor make out of themselves an exact copy of it. He/she can only describe this piece of knowledge with technology available to him/her by giving the determined significant state or the series of determined significant states to himself/herself or to some external object(s).

So, the word "description" becomes the key word, which requires a separate systematic review. Here, we will hold its examination in the "compressed" form.

7 Objects and their descriptions

The components of the knowledge embodied in the person at a time, at least, are: it's physical body and the ongoing thinking process in it (We are not talking about independent, but of interpenetrating components.). For a man to share his knowledge with the world means to "take" and to "display" some part of his thinking process outside.

Participants of the thinking process (here we are starting to build some information model of it) are, at least, the internal representations of various real and abstract objects that can be "rendered out" in two ways – in "analog" or "digital" way (The terms "analog" and "digital" are used as the nearest of the possible, which are capable to convey the idea.).

What we call here the "analog" way represents a high degree of similarity of the physical characteristics of the displayed and displaying objects. For example:

- in thought we have a three-dimensional spherical object (ball), which we can display either as a circle or a "filled" circle, where irregularity of "filling" simulates different levels of illumination of surface of the ball and, therefore, better conveys the idea of a three-dimensional object;
- a musical image can be the participant of the thinking process and can be displayed by voice or with help of some musical instrument;
- the idea of some flavor can be the participant of the thinking process and can be somehow accurately reproduced in the reality having the necessary ingredients and certain skills.

What we call "digital" way, is a refusal of the similarity between the physical characteristics of the displayed and displaying objects in favor of use of characters (digits, signs, codes, etc.). For example:

- the use of fixed grimaces, postures, sounds, about meaning of which there are the certain stable arrangements;
- the use of speech, the essence of which as the mapping instrument is stable relationship between abstract concepts and sound forms - about the same in Russian we say "шар", in German «Kugel», in English - «ball».

In this case, apparently, it is meaningless to search any similarities between the physical characteristics of the displayed and displaying objects. It's like trying to assess which of the options - Russian, German or English - better conveys the idea of the ball.

Here it seems appropriate to comment on the difference between the two known approaches to writing:

1) phonetic approach, in which the fixed signs are assigned to abstract objects of special kind - sounds. If acoustics is the science of sounds at all, then its sub-sections study, in particular, the "speech" sounds and the "music" sounds. And if speech sounds are displayed by signs of "writing", then for mapping of the music sounds the "phonetic"

approach uses some musical notation. The speech in a particular language displays the thinking process fragments, and, in turn, appears as the text.

2) hieroglyphic, in which the fixed signs (hieroglyphs) are assigned to abstract objects of any kind. This allows the direct display of the thinking process fragments in the "text".

In summary, we state the principle of distinguishing between objects and their descriptions. To the realization of this principle living creatures make their contribution – their activities are usually accompanied by appearance of various real objects, not present previously in Nature. In this case, in these processes of creating of new objects a two-step transition can be followed: the first step is to create a description of the target object, and the second - to implement new object in correspondence to the existing description.

An intermediate object – description, appearing in such process, is a necessary self-worth link of chain, without which the entire chain of activities (process) has no chance of success.

The description of the new object can appear in the thinking process for various reasons, the species of which we will not specifically examine. Speculative description may be in the same process of thought put to the test – is there an object that is needed, whether will it have the desired properties or not?

8 The role of consciousness in cognition and transformation of the World, as well as in communication between its carriers

Of course, we understand some extensibility of the concept of "consciousness" that we use. However, we are inclined to say that in all following cases:

- production of the description of the observed object;
- production of the description of the object to be created;
- production of the external description of the thinking process fragments
- consciousness plays a crucial role by sending the initial impulse and launching technologies available for this [7].

9 Information as an outcome of the process (definition)

The term "information" can be interpreted in two ways [6]:

- as the designation of process, which transforms something shapeless into some form;
- as a result of the "information" or "transformation" processes, when they are limited to product namely object description, but not the object itself.
- In this case, the "information" as a result (outcome) can be defined as the description of somewhat/someone consciously presented in the defined form (It seems that all concepts supporting this definition were disclosed with reasonable certainty).

10 Information interaction

Considering "information" we usually use a notion of the "information exchange", in which the supplier (source) of data S, the consumer (receiver) of data C, and the mediation (media) linking S and C, are distinguished. Usually the "information exchange" is considered

from different points of view but on foreground of consideration come such aspects of information as syntax, semantics and pragmatics of the informational message.

In order to begin to generate information a certain motivating impulse - internal or external - should act on S. Internal impulse can be represented as the desire/need/obligation to share with someone some part of own knowledge, more or less clearly identified (internal identifier).

External impulse - it is a request or an order to share some knowledge that comes to S outside, for example, from C. In this case, C has to identify somehow for S this needed knowledge (external identifier).

For its part, S may not be interested in the transfer of knowledge. He may refuse to grant it or may provide something different (disinformation).

Thus, we see that the "information exchange" is a very complex process, the success of which depends on many factors, which include, inter alia:

- the accuracy of identification (internal, external) of exchanging knowledge;
- the quality and degree of consistency of knowledge displaying (of S) and knowledge perception (of C) technologies;
- the degree of similarity of value scales of S and C;
- the degree of similarity of semantic fields of S and C.

11 Information and Life

We've already thought of such kind of machines as a computer, and viewed its relationship to "information". Next machine, which can be viewed in connection with the "information" is a machine called "Life". Many of us sincerely believe that the functioning of this machine is completely determined by the information code, the carrier of which is DNA.

However, we will try to argue. Certainty: the Nature has created a machine called "Life." One of its components is a causal mechanism for reproducing processes of a certain kind. Launched in action once, he lives and works. Or dies and doesn't exist. We can, on the basis of already conceived cause-effect relations, interfere in it, restructure it, and reproduce any desired processes. We can, without basing on anything, break it. We can, by intervening in it with a deficit of understanding, achieve the most unexpected consequences.

But all that is not so important. The main question is, how it started? That is, was it a momentary act, implemented in accordance with the laws of physics (fundamental interactions), or the appearance of Life was preceded by any "blueprints" - "information". If "yes", how did they look? Where are they now?

On the other hand, it's clear, that the shellac maps pre-existing sound object using a specific technology. This technology is reversible - the mirror image technology corresponds to it and allows obtain the new real sound object again, similar to the original according map (It is clear that quality of map will be inferior to the original, since it maps only a part of it.).

It is tempting to see in the DNA the map of someone/something (information). However, there is a nuance. A map in principle is always a partial image of the object and with the help of what technology and how this object will be recreated – it's to decide, ultimately, to consciousness. In the case of DNA it comes about the identity of the object - this is the object itself at a certain stage of its development cycle.

12 Conclusion

It seems that the definition of concept of "information" constructed here:

- meets our most common intuitive ideas about this phenomenon, in particular, does not go beyond his native Latin meaning (clarification, narration) and can serve as the starting point in determining our attitude to other approaches;
- allows to clearly define the subject of informatics as basic and applied discipline, taking away unacceptable to it the role of the All-Things-Science.

Attempts to invade in the field of physics of inert matter and living beings can certainly use some analogies, but it is clear that penetration into the mysteries of life requires diversification of the terms, search and discovery of such terms that could serve as tools to expand and deepen our vision and understanding of objective reality.

References

1. Capurro, Rafael & Hjørland, Birger (2003). The concept of information. Annual review of information science and technology (s. 343-411). Medford, N.J.: Information Today. A version retrieved November 6, 2011 from:http://www.capurro.de/infoconcept.html.
2. Sharma N. The Origin of the “Data Information Knowledge Wisdom” Hierarchy [Updated: February 4, 2008]. - http://erealityhome.wordpress.com/2008/03/09/the-origin-of-dikw-hierarchy/
3. Малышев О. В. Как «информация» становится «знанием»? – Системи підтримки прийняття рішень. Теорія і практика: Збірник доповідей науково-практичної конференції з міжнародною участю. – 9 червня 2008. – Київ: Інститут проблем математичних машин і систем НАН України. – 2008. – С. 147-150.
4. Малышев О.В. Воплощённое знание // Математичні машини і системи, 2009. - № 1. – С. 55-69.
5. Малышев О. В. Анти-DIKW. – Обчислювальний інтелект (результати, проблеми, дослідження): Матеріали 1-ї Міжнародної науково-технічної конференції (10-13 травня 2011 р. Черкаси). – Черкаси: Маклаут, 2011. – С. 104-105.
6. Малышев О. В. «Информация» как процесс и его результат. – Системи підтримки прийняття рішень. Теорія і практика: Збірник доповідей дистанційної науково-практичної конференції з міжнародною участю. – 6 червня 2012. – Київ: Інститут проблем математичних машин і систем НАН України. – 2012. – С. 119-122.
7. Малышев О.В. Информация как сознательный феномен // Математичні машини і системи, 2012. - № 3. – С. 166-177.

Investigation of the High Intensive Markov-Modulated Poisson Process

Alexander Moiseev[1], Anatoly Nazarov[2]

[1,2] Tomsk State University, Russian Federation
[1] alexander-moiseev@mail.ru, [2] nazarov@fpmk.tsu.ru

Abstract. Markov-Modulated Poisson Process (MMPP) with high rate is considered in the paper. It was shown that in condition of infinitely grown of the process rate probability distribution of event arrivals during any fixed time interval can be approximated by normal distribution. Parameters of the appropriate distribution are received in the paper.

Key words: random process, arrival process, Markov-Modulated Poisson Process

Consider Markov-Modulated Poisson Process [1]. Let its controlling Markov chain has K states, transitions of the chain are defined by infinitesimal matrix $N\mathbf{Q} = \{Nq_{vk}\}_{v,k=\overline{1,K}}$, where scalar N has a mean of great value (in theoretical research it is considered that $N \to +\infty$). Matrix **Q** has a property

$$q_{vv} = -\sum_{k \neq v} q_{vk}$$

or in matrix form:

$$\mathbf{QE} = \mathbf{0} \tag{1}$$

where **E** is column vector composed of 1, **0** is zero row vector.

Let conditional rate of the process at each state k is equal to $N\lambda_k$ ($k = \overline{1,K}$). Denote matrix of conditional rates as

$$N\mathbf{\Lambda} = \mathrm{diag}\{N\lambda_1, \ldots, N\lambda_K\}.$$

We have named this arrival process as *High Intensive Markov-Modulated Poisson Process* (or *HIMMPP*) because a great value N is present in its intensity and transition matrix.

Let denote a number of events in the process which arrives during time interval with length t by $m(t)$, and state of the controlling Markov chain at moment t by $k(t)$. Consider the Markovian two-dimensional random process $\{m(t), k(t)\}$. Put the following notation

$$P(m,k,t) = \mathbf{P}\{m(t) = m, k(t) = k\}.$$

For this function we can write the following equation by using full probability formula:

$$P(m,k,t+\Delta t) = P(m,k,t)\cdot(1 - N\lambda_k \Delta t)\cdot(1 + Nq_{kk}\Delta t) + \\ + P(m-1,k,t)N\lambda_k \Delta t + \sum_{\nu \neq k} P(m,\nu,t)Nq_{\nu k}\Delta t + o(\Delta t)$$

or

$$P(m,k,t+\Delta t) - P(m,k,t) = \left[-P(m,k,t) + P(m-1,k,t)\right]N\lambda_k \Delta t + \\ + \sum_{\nu=1}^{K} P(m,\nu,t)Nq_{\nu k}\Delta t + o(\Delta t).$$

From this expression in condition of $\Delta t \to 0$ we obtain Kolmogorov equation:

$$\frac{1}{N}\frac{\partial P(m,k,t)}{\partial t} = \left[-P(m,k,t) + P(m-1,k,t)\right]N\lambda_k + \sum_{\nu=1}^{K} P(m,\nu,t)Nq_{\nu k}\ .$$

Let multiply left and right parts of this equation by e^{jum}, where $j = \sqrt{-1}$ and u is some variable. Then sum it over m from 0 up to ∞. So, using notation

$$H(u,k,t) = \sum_{m=0}^{\infty} e^{jum} P(m,k,t)$$

we obtain

$$\frac{1}{N}\frac{\partial H(u,k,t)}{\partial t} = H(u,k,t)\lambda_k\left(e^{ju} - 1\right) + \sum_{\nu=1}^{K} H(u,k,t)q_{\nu k}\ . \qquad (2)$$

Let row vector $\mathbf{H}(u,t)$ is $\mathbf{H}(u,t) = \{H(u,1,t),\ldots,H(u,K,t)\}$, then we can rewrite (2) in the following form:

$$\frac{1}{N}\frac{\partial \mathbf{H}(u,t)}{\partial t} = \mathbf{H}(u,t)\mathbf{\Lambda}\left(e^{ju} - 1\right) + \mathbf{H}(u,t)\mathbf{Q} = \mathbf{H}(u,t)\left[\mathbf{Q} + \mathbf{\Lambda}\left(e^{ju} - 1\right)\right].$$

In this equation we make a substitution

$$\mathbf{H}(u,t) = \mathbf{H}_2(u,t)e^{juN\lambda t},$$

where

$$\lambda = \mathbf{R}\Lambda\mathbf{E}, \tag{3}$$

R is row vector of stationary distribution for states of the controlling Markov chain. This vector satisfies to the conditions:

$$\begin{cases} \mathbf{RQ} = \mathbf{0}, \\ \mathbf{RE} = 1. \end{cases} \tag{4}$$

As a result we obtain an equation for function $\mathbf{H}_2(u,t)$:

$$\frac{1}{N}\frac{\partial \mathbf{H}_2(u,t)}{\partial t} e^{juN\lambda t} + ju\lambda \mathbf{H}_2(u,t) e^{juN\lambda t} = \mathbf{H}_2(u,t)\left[\mathbf{Q} + \Lambda\left(e^{ju} - 1\right)\right] e^{juN\lambda t}$$

or

$$\frac{1}{N}\frac{\partial \mathbf{H}_2(u,t)}{\partial t} = \mathbf{H}_2(u,t)\left[\mathbf{Q} + \Lambda\left(e^{ju} - 1\right) - ju\lambda\mathbf{I}\right], \tag{5}$$

where **I** is identity matrix of size *K*.

This equation we will solve by the asymptotic analysis method [1]. We use notation $\varepsilon^2 = \frac{1}{N}$ and make substitutions $u = \varepsilon w$ и $\mathbf{H}_2(u,t) = \mathbf{F}(w,t,\varepsilon)$. So equation (5) will be rewritten in a form:

$$\varepsilon^2 \frac{\partial \mathbf{F}(w,t,\varepsilon)}{\partial t} = \mathbf{F}(w,t,\varepsilon)\left[\mathbf{Q} + \Lambda\left(e^{j\varepsilon w} - 1\right) - j\varepsilon w\lambda\mathbf{I}\right]. \tag{6}$$

Designate by

$$\mathbf{F}(w,t) = \lim_{\varepsilon \to 0} \mathbf{F}(w,t,\varepsilon).$$

Prove the following statement.

Theorem. The solution of the equation (6) has the following form in condition of $\varepsilon \to 0$:

$$\mathbf{F}(w,t) = \mathbf{R}\exp\left\{\frac{(jw)^2}{2}\kappa t\right\},$$

where

$$\kappa = \lambda + \mathbf{f}(\Lambda - \lambda\mathbf{I})\mathbf{E}, \tag{7}$$

and row vector **f** satisfies to equation

$$\mathbf{R}(\mathbf{\Lambda} - \lambda\mathbf{I}) + \mathbf{fQ} = \mathbf{0}.$$

Proof. We will make the proof at three stages.

Stage 1. Let $\varepsilon \to 0$ in (6), we get:

$$\mathbf{F}(w,t)\mathbf{Q} = \mathbf{0}.$$

This equation has a same form as first condition in the (4), so vector function $\mathbf{F}(w, t)$ can be written in form

$$\mathbf{F}(w,t) = \mathbf{R}\Phi(w,t), \tag{8}$$

where $\Phi(w,t)$ is some scalar function.

Stage 2. We will search a solution of the equation (6) in expansion form

$$\mathbf{F}(w,t,\varepsilon) = \Phi(w,t)\left[\mathbf{R} + j\varepsilon w\mathbf{f}\right] + \mathbf{O}(\varepsilon^2), \tag{9}$$

where **f** is some row vector, $\mathbf{O}(\varepsilon^2)$ is row vector which consists of infinitesimal values of the order ε^2. Substituting this expression into (6) and using an approximation $e^{j\varepsilon w} = 1 + j\varepsilon w + O(\varepsilon^2)$ we get:

$$\varepsilon^2\left[\mathbf{R} + j\varepsilon w\mathbf{f}\right]\frac{\partial\Phi(w,t)}{\partial t} = \Phi(w,t)\left[\mathbf{R} + j\varepsilon w\mathbf{f}\right]\left[\mathbf{Q} + j\varepsilon w\mathbf{\Lambda} - j\varepsilon w\lambda\mathbf{I}\right] + \mathbf{O}(\varepsilon^2)$$

.

From where in condition of $\varepsilon \to 0$ we get the following equation for vector **f**:

$$\mathbf{R}(\mathbf{\Lambda} - \lambda\mathbf{I}) + \mathbf{fQ} = \mathbf{0}.$$

Stage 3. Let sum the equations at the left and at the right parts of the system (6). To do this we make a right multiplication of each part of (6) on identity column vector **E** with length *K*:

$$\varepsilon^2\frac{\partial\mathbf{F}(w,t,\varepsilon)}{\partial t}\mathbf{E} = \mathbf{F}(w,t,\varepsilon)\left[\mathbf{Q} + \mathbf{\Lambda}\left(e^{j\varepsilon w} - 1\right) - j\varepsilon w\lambda\mathbf{I}\right]\mathbf{E}.$$

Using an approximation $e^{j\varepsilon w} = 1 + j\varepsilon w + \frac{(j\varepsilon w)^2}{2} + O(\varepsilon^3)$ in this equation and taking into account (1) we get:

$$\varepsilon^2 \frac{\partial \mathbf{F}(w,t,\varepsilon)}{\partial t}\mathbf{E} = \mathbf{F}(w,t,\varepsilon)\left[j\varepsilon w(\mathbf{\Lambda} - \lambda\mathbf{I}) + \frac{(j\varepsilon w)^2}{2}\mathbf{\Lambda}\right]\mathbf{E} + O(\varepsilon^3)\,.$$

Substitute (9) in this equation:

$$\varepsilon^2 \frac{\partial \Phi(w,t)}{\partial t}\mathbf{RE} = \Phi(w,t)\left[j\varepsilon w\mathbf{R}(\mathbf{\Lambda} - \lambda\mathbf{I}) + \frac{(j\varepsilon w)^2}{2}\mathbf{R\Lambda} + (j\varepsilon w)^2\mathbf{f}(\mathbf{\Lambda} - \lambda\mathbf{I})\right]\mathbf{E} + \mathbf{O}(\varepsilon^3).$$

Applying (3–4) and making simple transforms we get:

$$\frac{\partial \Phi(w,t)}{\partial t} = \frac{(jw)^2}{2}\Phi(w,t)\left[\lambda + \mathbf{f}(\mathbf{\Lambda} - \lambda\mathbf{I})\mathbf{E}\right] + O(\varepsilon)\,.$$

In condition of $\varepsilon \to 0$ we get differential equation on unknown scalar function $\Phi(w,t)$:

$$\frac{\partial \Phi(w,t)}{\partial t} = \frac{(jw)^2}{2}\kappa\Phi(w,t)\,,$$

where value of κ is defined by expression (7). Boundary condition for this equation has a form $\Phi(w,0) = 1$ because it is obvious that

$$P(m,0) = \begin{cases} \mathbf{R} \text{ if } m = 0, \\ 0 \text{ if } m > 0. \end{cases}$$

This boundary value problem has a solution

$$\Phi(w,t) = \exp\left\{\frac{(jw)^2}{2}\kappa t\right\},$$

from where we get:

$$\mathbf{F}(w,t) = \mathbf{R}\exp\left\{\frac{(jw)^2}{2}\kappa t\right\}$$

due to (8). The theorem is proved.

Returning to function $\mathbf{H}(u,t)$ in condition of N is great enough we get approximation

$$\mathbf{H}(u,t) \approx \mathbf{R}\exp\left\{ juN\lambda t + \frac{(ju)^2}{2}N\kappa t\right\}.$$

Therefore characteristic function

$$h(u,t) = \mathbf{H}(u,t)\mathbf{E}$$

of the process $m(t)$ has a form of characteristic function for the random variable with normal probability distribution. So, a high intensive MMPP in condition of $N \to \infty$ has the same characteristic function as a Gaussian process. That is a probability distribution of event arrivals in time interval t for HIMMPP may be approximated by normal probability distribution with expected value $N\lambda t$ and variance $N\kappa t$. The similar results were obtained for some other types of high intensive processes: MAP, general independent [2].

References

1. Nazarov, A.A., Moiseeva, S.P.: The Asymptotical Analysis Method in Queueing Theory. Tomsk, NTL (2006)
2. Moiseev A., Nazarov A.: Investigation of High Intensive General Flow // Proceedings of the IV International Conference "Problems of Cybernetics and Informatics" (PCI'2012), September 12-14, 2012 Baku, Azerbaijan, pp. 161–163.

Modeling of Insurance Company as Infinite-Servers Queueing System

Alexander Moiseev[1], Svetlana Moiseeva[2], Irina Sinyakova[3]

[1,2,3] Tomsk State University, Russian Federation

[1] alexander-moiseev@mail.ru, [2] smoiseeva@mail.ru, [3] irinka_asf@mail.ru

Abstract. Mathematical model for insurance company in a form of M|M|∞ queueing system is considering in the paper. Analytical expression for the moment-generating function, mean and variance of insurance company funds are obtained.

Key words: queueing theory, infinite-servers model, mathematical modeling for insurance

Introduction

Many of human activities lead to risk situations. In most cases consequences of the situations are financial. The insurance reduces a risk of financial losses. To achieve that insurance companies collect insurance payments from independent sources, and they can be ready to cover financial losses of any customer.

So, basic principle of insurance consists in the following. Insurance company (*insurer*) initially takes from a customer (*insured*) a payment which is named *insurance premium*. Later, if insured event occurs the insurer must make a payment of the insurance (*insurance payment*). Insurer can periodically pay additional payments during insurance period, we will name their *insurance fees*.

We assume that values of insurance premiums and insurance payments are random values. Time moments of premiums income and payments outcome are considered as random variables too. So an insurance company activity can be regarded as mathematical model based on random variables and processes. In the paper a market of insurance services is considered as infinite [1].

The paper objective is analysis of statistic characteristics of two random processes: $i(t)$ – a number of insured at time moment t, and $S(t)$ – value of company funds.

Mathematical Model

Consider a model of insurance company in a form of an infinite-servers queueing system M|M|∞ with Poisson process input [2]. Denote the rate of input process by λ. Incoming customer occupies any free server and stays there during random time distributed by exponential law with parameter μ.

So, $i(t)$ will be a number of busy servers. Denote a number of customers that arrive during time interval t by $l(t)$. Capital of the company is formed by insurance premiums and insurance fees and decreases by insurance payments.

Put the following notations:

- φ – value of an insurance premium, let it be a random variable with distribution function $F_{\varphi}(x)$; denote its moments as $a_1 = \mathrm{M}\{\varphi\}$ and $a_2 = \mathrm{M}\{\varphi^2\}$;
- ξ – value of an insurance fee, let it be a random variable with distribution function $F_{\xi}(x)$ and moments $b_1 = \mathrm{M}\{\xi\}$, $b_2 = \mathrm{M}\{\xi^2\}$;
- η – value of an insurance payment, let it be a random variable with distribution function $F_{\eta}(x)$ and moments $c_1 = M\{\eta\}$, $c_2 = M\{\eta^2\}$;
- ψ_{η} - a rate of insured event arrivals.

We can write an expression for an insurance company capital:

$$S(t) = \sum_{l=1}^{l(t)} \varphi_l + \sum_{i=1}^{i(t)} (\xi_i - \psi_{\eta}\eta_i).$$

Research of the two-dimensional process included a number of customers presented in the system and a number of customer arrivals

Designate probability distribution for the two-dimensional Markov chain which consists of a number of customers presented in the system at the moment t and a number of customer arrivals during time period t by $P(l,i,t) = \mathrm{P}\{l(t) = l, i(t) = i\}$, $l = \overline{0,\infty}$ and $i = \overline{0,\infty}$.

We construct Kolmogorov differential equations system for the probability $P(l,i,t)$ by the Δt-method. Using formula of total probability we get:

$$P(l,i,t+\Delta t) = (1-\lambda\Delta t)(1-i\mu\Delta t)P(l,i,t) + \lambda\Delta t P(l-1,i-1,t) + \\ + (i+1)\mu\Delta t P(l,i+1,t) + o(\Delta t),$$

from where we obtain the system

$$\frac{\partial P(l,i,t)}{\partial t} = -(\lambda + i\mu)P(l,i,t) + \lambda P(l-1,i-1,t) + (i+1)\mu P(l,i+1,t). \qquad (10)$$

Boundary condition will be the following:

$$P(l,i,0) = \begin{cases} 1, & \text{if } i = l = 0, \\ 0, & \text{otherwise.} \end{cases}$$

We write a moment-generating function for the distribution $P(l,i,t)$ in a form

$$F(x,y,t)=\sum_{l=0}^{\infty}\sum_{i=0}^{\infty}x^l y^i P(l,i,t).$$

From the system (1) we obtain the following:

$$\frac{\partial F(x,y,t)}{\partial t}=-\lambda F(x,y,t)-\mu\sum_{l=1}^{\infty}\sum_{i=0}^{\infty}iy^i x^l P(l,i,t)+$$
$$+\lambda\sum_{l=1}^{\infty}\sum_{i=1}^{\infty}x^l y^i P(l-1,i-1,t)+\mu\sum_{l=0}^{\infty}\sum_{i=0}^{\infty}(i+1)y^i x^l P(l,i+1,t),$$

from where we get a first-order linear differential equation in partial derivatives for the function $F(x,y,t)$:

$$\frac{\partial F(x,y,t)}{\partial t}=-\lambda F(x,y,t)-\mu y\frac{F(x,y,t)}{\partial y}+\lambda xyF(x,y,t)+\mu\frac{F(x,y,t)}{\partial y}.$$

So we get the following:

$$\frac{\partial F(x,y,t)}{\partial t}+\mu(y-1)\frac{\partial F(x,y,t)}{\partial y}=\lambda(xy-1)F(x,y,t). \qquad (11)$$

We can construct the following ordinary differential equation system for characteristics of the equation (2) [3]:

$$\frac{dt}{1}=\frac{dy}{\mu(y-1)}=\frac{dF(x,y,t)}{\lambda(xy-1)F(x,y,t)}.$$

Rewrite this system as

$$\frac{dt}{1}=\frac{dy}{\mu(y-1)}=\frac{dF(x,y,t)}{\lambda\left[(x-1)(y-1)+(x-1)+(y-1)\right]F(x,y,t)}.$$

The first integral of this system equals

$$C_1=(y-1)e^{-\mu t}. \qquad (12)$$

The second integral equals

$$F(x,y,t)=C_2\cdot\exp\left\{\frac{\lambda}{\mu}(x-1)\cdot C_1e^{\mu t}+\lambda(x-1)t+\frac{\lambda}{\mu}C_1e^{\mu t}\right\}. \qquad (13)$$

Substituting expression (3) into expression (4) we can write

$$F(x,y,t)=\Phi\left((y-1)e^{-\mu t}\right)\cdot\exp\left\{\frac{\lambda}{\mu}(x-1)(y-1)+\lambda(x-1)t+\frac{\lambda}{\mu}(y-1)\right\}, \qquad (14)$$

where $\Phi(x)$ is some function. Using boundary conditions we obtain

$$F(x,y,0)=\Phi(y-1)\cdot\exp\left\{\frac{\lambda}{\mu}(x-1)(y-1)+\frac{\lambda}{\mu}(y-1)\right\}=1,$$

and so

$$\Phi\left((y-1)e^{-\mu t}\right)=\exp\left\{-\frac{\lambda(x-1)(y-1)}{\mu}e^{-\mu t}-\frac{\lambda}{\mu}(y-1)e^{-\mu t}\right\}.$$

Substituting this into (5) we obtain the following expression for the moment-generating function $F(x,y,t)$:

$$F(x,y,t)=\exp\left\{\frac{\lambda}{\mu}(x-1)(y-1)\left(1-e^{-\mu t}\right)+\frac{\lambda}{\mu}(y-1)\left(1-e^{-\mu t}\right)+\lambda(x-1)t\right\}. \qquad (15)$$

Basic probability characteristics for insurance company funds

Consider a characteristic function for the insurance company capital value accumulated during period *t*:

$$H(\alpha,t)=\mathrm{M}\left\{e^{\alpha S(t)}\right\}=\mathrm{M}\left\{e^{\alpha\left[\sum_{l=1}^{l(t)}\varphi_l+\sum_{i=1}^{i(t)}\left(\xi_i-\psi_\eta\eta_i\right)\right]}\right\}=$$

$$=\sum_{l=0}^{\infty}\sum_{i=0}^{\infty}\mathrm{M}\left\{e^{\alpha\left[\sum_{l=1}^{l(t)}\varphi_l+\sum_{i=1}^{i(t)}\left(\xi_i-\psi_\eta\eta_i\right)\right]}\middle| l(t)=l,i(t)=i\right\}P(l,i,t)=$$

$$=\sum_{l=0}^{\infty}\sum_{i=0}^{\infty}\left(\mathrm{M}e^{\alpha\varphi_l}\right)^l\cdot\left(\mathrm{M}e^{\alpha\left(\xi_i-\psi_\eta\eta_i\right)}\right)^i P(l,i,t)=\sum_{l=0}^{\infty}\sum_{i=0}^{\infty}\varphi(\alpha)^l\cdot\psi(\alpha)^i P(l,i,t).$$

Here $\varphi(\alpha)=\mathrm{M}e^{\alpha\varphi_l}$ is a characteristic function for random variable φ_l, $\psi(\alpha)=\mathrm{M}e^{\alpha\left(\xi_i-\psi_\eta\eta_i\right)}$ is a characteristic function for random variable $\xi_i-\psi_\eta\eta_i$.

Using an expression for the generating function (6) we can write the following:

$$H(\alpha,t)=\mathrm{M}e^{\alpha S(t)}=F(\varphi(\alpha),\psi(\alpha),t)=$$

$$=\exp\left\{\frac{\lambda}{\mu}(\varphi(\alpha)-1)(\psi(\alpha)-1)\left(1-e^{-\mu t}\right)+\frac{\lambda}{\mu}(\psi(\alpha)-1)\left(1-e^{-\mu t}\right)+\lambda(\varphi(\alpha)-1)t\right\}.$$

Mean of the insurance company capital value accumulated during period *t* equals

$$\mathrm{M}\,S(t)=\left.\frac{\partial H(\alpha,t)}{\partial\alpha}\right|_{\alpha=0}$$

and

$$\varphi(\alpha)\big|_{\alpha=0}=1,$$

$$\varphi'(\alpha)\big|_{\alpha=0}=\mathrm{M}\left\{\varphi_l e^{\alpha\varphi_l}\right\}\Big|_{\alpha=0}=\mathrm{M}\{\varphi_l\}=a_1,$$

$$\varphi''(\alpha)=\mathrm{M}\{\varphi_l^2\}=a_2.$$

And also

$$\psi(\alpha)\big|_{\alpha=0}=1,$$

$$\psi'(\alpha)\big|_{\alpha=0}=\mathrm{M}\left\{(\xi_i-\psi_\eta\eta_i)e^{(\xi_i-\psi_\eta\eta_i)}\right\}\Big|_{\alpha=0}=\mathrm{M}\{\xi_i\}-\psi_\eta\,\mathrm{M}\{\eta_i\}=b_1-\psi_\eta\cdot c_1,$$

$$\psi''(\alpha)\big|_{\alpha=0}=\mathrm{M}\{\xi_i^2\}-\psi_\eta^2\,\mathrm{M}\{\eta_i^2\}=b_2-\psi_\eta^2\cdot c_2.$$

So the mean (the first moment) of the insurance company capital value accumulated during period *t* is

$$\mathrm{M}\,S(t)=\left\{\frac{\lambda}{\mu}(\varphi(\alpha)-1)(\psi(\alpha)-1)(1-e^{-\mu t})+\frac{\lambda}{\mu}(\psi(\alpha)-1)(1-e^{-\mu t})+\lambda(\varphi(\alpha)-1)t\right\}'\times$$

$$\times\exp\left\{\frac{\lambda}{\mu}(\varphi(\alpha)-1)(\psi(\alpha)-1)(1-e^{-\mu t})+\frac{\lambda}{\mu}(\psi(\alpha)-1)(1-e^{-\mu t})+\lambda(\varphi(\alpha)-1)t\right\}=$$

$$=\lambda t a_1+\frac{\lambda}{\mu}(1-e^{-\mu t})(b_1-\psi_\eta c_1).$$

The second moment can be obtained similarly:

$$\mathrm{M}\,S^2(t)=\left.\frac{\partial^2 H(\alpha,t)}{\partial\alpha^2}\right|_{\alpha=0}=$$

$$=\left[\left\{\frac{\lambda}{\mu}(\varphi(\alpha)-1)(\psi(\alpha)-1)(1-e^{-\mu t})+\frac{\lambda}{\mu}(\psi(\alpha)-1)(1-e^{-\mu t})+\lambda(\varphi(\alpha)-1)t\right\}''\times\right.$$

$$\times\exp\left\{\frac{\lambda}{\mu}(\varphi(\alpha)-1)(\psi(\alpha)-1)(1-e^{-\mu t})+\frac{\lambda}{\mu}(\psi(\alpha)-1)(1-e^{-\mu t})+\lambda(\varphi(\alpha)-1)t\right\}+$$

$$+\left\{\frac{\lambda}{\mu}(\varphi(\alpha)-1)(\psi(\alpha)-1)(1-e^{-\mu t})+\frac{\lambda}{\mu}(\psi(\alpha)-1)(1-e^{-\mu t})+\lambda(\varphi(\alpha)-1)t\right\}'^2\times$$

$$\left.\times\exp\left\{\frac{\lambda}{\mu}(\varphi(\alpha)-1)(\psi(\alpha)-1)(1-e^{-\mu t})+\frac{\lambda}{\mu}(\psi(\alpha)-1)(1-e^{-\mu t})+\lambda(\varphi(\alpha)-1)t\right\}\right]\Big|_{\alpha=0}=$$

$$=2\left[\frac{\lambda}{\mu}(1-e^{-\mu t})a_1c_1\right]+\frac{\lambda}{\mu}(1-e^{-\mu t})(b_2-\psi_\eta^2c_2)+\lambda t a_2+(\lambda t a_1)^2+\left[\frac{\lambda}{\mu}(1-e^{-\mu t})(b_1-\psi_\eta c_1)\right]^2$$

.

And we obtain that a variance of the insurance company capital value is equal to

$$\mathrm{D}\,S(t) = 2\left[\frac{\lambda}{\mu}\left(1-e^{-\mu t}\right)a_1c_1\right] + \frac{\lambda}{\mu}\left(1-e^{-\mu t}\right)\left(b_2-\psi_\eta^2 c_2\right) + \lambda t a_2 + \left(\lambda t a_1\right)^2 +$$

$$+\left[\frac{\lambda}{\mu}\left(1-e^{-\mu t}\right)\left(b_1-\psi_\eta c_1\right)\right]^2 - \left(\lambda t a_1\right)^2 - \left[\frac{\lambda}{\mu}\left(1-e^{-\mu t}\right)\left(b_1-\psi_\eta c_1\right)\right]^2 =$$

$$= \lambda t a_2 + \frac{\lambda}{\mu}\left(1-e^{-\mu t}\right)\left[2a_1c_1 + b_2 - \psi_\eta^2 c_2\right]$$

Conclusion

In the paper mathematical model of the insurance company activity was constructed in a form of M|M|∞ queueing system. Research of two-dimensional system states process gave an analytical expression for characteristic function of the insurance company capital value. In consequence of that basic probability characteristics (mean and variance) for the insurance company funds were obtained.

References

1. Glukhova E.V., Zmeyev O.A., Livshits K.I.: Mathematical models of insurance. Tomsk, Tomsk University (2004)
2. Nazarov, A.A., Terpugov A.F.: Queueing Theory. Tomsk, NTL (2010)
3. Elsgolts L.E.: Differential Equations and Calculus of Variations. Moscow, Science (1969)

Asymptotical Research of the Output Process of BMAP|M|∞ queue

Ivan Lapatin, Svetlana Lopukhova

Tomsk State University, Russian Federation
ilapatin@mail.ru, lopuchovasv@mail.ru

Abstract. In the paper the output process of infinite-server queue with batch Markovian arrival process (BMAP) is considered. Asymptotical approximation of the characteristic function of the customers number served during some time period is obtained. The output process of BMAP|M|∞ queue in condition of rising service time was shown to be asymptotical Poisson.

Key words: output process, BMAP, infinite-servers model.

The study of the output processes in queueing systems is primarily motivated by the need to analyze queueing network models, in which the output process of one queue is the arrival process of another queue. Main results in the investigation of the output processes of the classic models were obtained by Burke P. [1], Reich E. [2] and Finch P. [3] in the second half of the XX century. They showed that the output process of a queue with Poisson input and an exponential service is Poisson with the same rate as the arrivals. In 1963 Mirasol N. proved that the output of the M|G|∞ is Poisson too. But departure process characterization of a queueing system has been difficult unless the queue has very special structure.

In this work the output process of an infinite-server queue with batch Markovian arrival process (BMAP) is considered. Arrival process is set by the matrix of Markov's control chain $k(t)$ infinitesimal characteristics $\mathbf{Q}$, a set of conditional rates λ_k ($k = 1,\ldots,K$) and a set of probabilities $p_{k\nu}(s)$ (k, $\nu=1,\ldots,K$) for different integer values s. Special cases of BMAPs include Poisson processes, phase-type (PH) renewal processes, Markov-modulated Poisson processes (MMPPs), and Markovian arrival processes (MAPs). First definition of the BMAP was given by Lucantoni D [5]. Call arrival occupies any channel and is served during the random time. Service times are independent and identically distributed by exponential law random variables.

The output process of this system is investigated by the method of asymptotical analysis [6] in condition of rising service time. This condition denotes that mean of the service time converges to infinity.

Denote the process $m(t)$ – the number of customers that have completed service in the system at time t, $i(t)$ – the number of customers in the service of the system at time t. Then the process $\{k(t), m(t), i(t)\}$ is a Markov process. For the probability distribution
$P\{k(t)= k, m(t)= m, i(t)=i\}$
we will receive a system of Kolmogorov differential equations.

$$\frac{\partial P(k,i,m,t)}{\partial t} = -(\lambda_k + i\mu)P(k,i,m,t) + q_{kk}P(k,i,m,t) +$$
$$+ (i+1)\mu P(k,i+1,m-1,t) + \sum_{s=1}^{i} \lambda_k p_{kk}(s) P(k,i-s,m,t) +$$
$$+ \sum_{s=0}^{i} \sum_{v \neq k}^{K} q_{vk} p_{vk}(s) P(v,i-s,m,t).$$

We write the $P(k,i,m,t)$ as a row vector

$$\mathbf{P}(i,m,t) = \{P(1,i,m,t), P(2,i,m,t), P(3,i,m,t), \dots\}$$

and introduce the matrix **d**(s) with the elements

$$d_{vk}(s) = \begin{cases} \lambda_k p_{kk}(s), & s = \overline{1,\infty}, \\ q_{vk} p_{vk}(s), & s = \overline{0,\infty}. \end{cases}$$

We denote $p_{kk}(0) = 0$, then we get a system of Kolmogorov differential equations

$$\frac{\partial \mathbf{P}(i,m,t)}{\partial t} = -i\mu \mathbf{P}(i,m,t) + (i+1)\mu \mathbf{P}(i+1,m-1,t) + \sum_{s=0}^{i} \mathbf{d}(s)\mathbf{P}(i-s,m,t). \quad (16)$$

We denote the function

$$\mathbf{H}(x,u,t) = \sum_{i=0}^{\infty} e^{jxi} \sum_{m=0}^{\infty} e^{jum} \mathbf{P}(i,m,t), \quad (17)$$

which will be called a function similar to characteristic, where $j = \sqrt{-1}$ – the imaginary unit.

For these functions of the system (1) the following Cauchy problem can be written

$$\frac{\partial \mathbf{H}(x,u,t)}{\partial t} + j\mu(e^{ju}e^{-jx} - 1)\frac{\partial \mathbf{H}(x,u,t)}{\partial x} = \mathbf{H}(x,u,t)\mathbf{D}(x), \quad (18)$$

where $\mathbf{D}(x) = \sum_{s=0}^{\infty} e^{jsx}\mathbf{d}(s)$.

We let $\mu \to 0$, we denote $\mu = \varepsilon$ and perform replacements in (3)

$x = \varepsilon y$, $\mathbf{H}(x, u, t) = \mathbf{F}(y, u, t, \varepsilon)$, (19)

and get the following Cauchy problem

$$\frac{\partial \mathbf{F}(y,u,t,\varepsilon)}{\partial t} + j(e^{ju}e^{-j\varepsilon y} - 1)\frac{\partial \mathbf{F}(y,u,t,\varepsilon)}{\partial y} = \mathbf{F}(y,u,t,\varepsilon)\mathbf{D}(\varepsilon y). \quad (20)$$

Theorem 1. Sum of the components of the limit values, by $\varepsilon \to 0$, of the row vector **F**(*y*, *u*, *t*), the solution **F**(*y*, *u*, *t*, ε) of equation (5) has the form

$$\mathbf{F}(y,u,t)\mathbf{E} = \exp\{jy\kappa + (e^{ju} - 1)\kappa t\}, \quad (21)$$

where **E** – the single column vector, and value κ is defined by equality

$$\kappa = \mathbf{R}\mathbf{D}_1\mathbf{E}, \qquad (22)$$

where

$$\mathbf{D}_1 = \sum_{s=0}^{\infty} s\mathbf{d}(s),$$

and the row vector **R** is defined by the system

$$\begin{cases} \mathbf{R}\mathbf{Q} = 0, \\ \mathbf{R}\mathbf{E} = 1. \end{cases} \qquad (23)$$

Proof. We multiply the equation (5) in the right on a single column vector **E** and perform the limit transition as $\varepsilon \to 0$

$$\frac{\partial \mathbf{F}(y,u,t)}{\partial t}\mathbf{E} + j(e^{ju} - 1)\frac{\partial \mathbf{F}(y,u,t)}{\partial y}\mathbf{E} = 0. \qquad (24)$$

The general solution of this equation is

$$\mathbf{F}(y,u,t)\mathbf{E} = \varphi\left(t + \frac{jy}{e^{ju} - 1}\right),$$

where $\varphi(y)$ is some function. We need an initial condition. Consider function $\mathbf{F}(y, u, t)\mathbf{E}$ at time 0. It is evident that $\mathbf{F}(y,u,0)\mathbf{E}$ does not depend on u (we simply begin to observe the output process at the moment of time equals zero then the number of the served customers is also equal zero).

Thus, we obtain the initial condition

$$\mathbf{F}(y,u,0)\mathbf{E} = \Phi(y). \qquad (25)$$

$\Phi(y)$ is an asymptotic approximation of the characteristic function of a stationary distribution of the number of servers employed in the system in the condition of rising service time

$$\Phi(y) = e^{jy\kappa},$$

where value κ is defined by (7). Value κ is the rate of input BMAP.

Thus, we can write the solution of the equation (9) as following

$$\mathbf{F}(y,u,t)\mathbf{E} = \exp\left\{jy\kappa + (e^{ju} - 1)\kappa t\right\}.$$

The main result of this paper is an asymptotical approximation of the characteristic function of the number of the customers $m(t)$ served during the time t

$$\begin{aligned} M\{e^{jum(t)}\} &= \mathbf{H}(0,u,t)\mathbf{E} = \mathbf{F}(0,u,t,\varepsilon)\mathbf{E} \approx \\ &\approx \mathbf{F}(0,u,t)\mathbf{E} = \exp\left\{(e^{ju} - 1)\kappa t\right\}, \end{aligned} \qquad (26)$$

where κ is the rate of BMAP.

Thus, the output process of BMAP|M|∞ queue in the condition of rising service time was shown to be asymptotical Poisson. The obtained results summarize the results for the model with MAP [7] in case of non-ordinary arrival process.

References

1. Burke P.J.: The Output of Queueing Systems. Operations Research, vol. 4, pp. 699–704 (1956)
2. Reich E.: Waiting Times When Queues are in Tandem. Ann. Math. Statist., vol. 28, No. 3, p. 768 (1957)
3. Finch P.D.: The Output Process of the Queueing System M|G|1. J. Roy. Statist. Soc., vol. 21, No. 2, pp. 375–380 (1959)
4. Mirasol N.M.: The output of an $M \mid G \mid \infty$ queueing system is Poisson. Operations Research, No. 11, pp. 282–284 (1963)
5. Lucantoni D.: New results for the single server queue with a batch Markovian arrival process. Stochastic Models, vol. 7, pp. 1–46 (1991)
6. Nazarov, A.A., Moiseeva, S.P.: The Asymptotical Analysis Method in Queueing Theory. Tomsk, NTL (2006)
7. Lapatin, I.L., Nazarov, A.A.: Reseach of $MAP \mid M \mid \infty$ Output in Condition of Rising Service Time. Probability Theory, Mathematical Statistics and Applications. Minsk, RIVS, vol. 2, pp. 76–80 (2009)

MODEL SUPPORT FOR INFORMATION SEARCH AT THE ONTOLOGY-ORIENTED PORTAL

A. F. Verlan[1], O. Oletsky[2]

[1] Institute for Simulation Problems in Power Engineering, Kyiv, Ukraine
a.f.verlan@gmail.com

[2] National University "Kiev-Mohyla Academy", Kyiv, Ukraine
oletsky@ukma.kiev.ua

Abstract. Organization of model support for information search at an ontology-oriented portal is regarded. An approach related to integration of the model "ontology-document" into the network of connected models.

Keywords: information search, ontology, relevance, model support

ОБ ОРГАНИЗАЦИИ МОДЕЛЬНОЙ ПОДДЕРЖКИ ИНФОРМАЦИОННОГО ПОИСКА В РАМКАХ ОНТОЛОГИЧЕСКИ-ОРИЕНТИРОВАННОГО ТЕМАТИЧЕСКОГО ПОРТАЛА

А. Ф. Верлань[1], Олецкий А. В.[2]

[1]Институт проблем моделирования в энергетике им. Г.Е.Пухова, Киев, Украина
a.f.verlan@gmail.com

[2]Национальный Университет «Киево-Могилянская Академия», Киев, Украина
oletsky@ukma.kiev.ua

Абстракт. Рассматриваются возможности, связанные с организацией модельной поддержки информационного поиска на тематическом портале. Предлагается подход, связанный с погружением графа "онтология-документ" в сеть взаимосвязанных моделей.

Ключевые слова: информационный поиск, онтология, релевантность, модельная поддержка

Введение

Проблема поиска информационных ресурсов, которые наиболее точно соответствуют целям и информационным потребностям пользователей, очень важна и актуальна, но она далека от полного решения. Существующие подходы [1-3 и др.]

носят в основном эвристический характер и затрагивают только отдельные аспекты проблемы.

Не вызывает сомнений, что повышение качества поиска, полноты и релевантности его результатов должно быть тесно связано с максимальной ориентацией на семантику, онтологию предметной области. Такая онтологическая ориентированность имеет особое значение для тематически однородных веб-ресурсов, для которых характерны высокая информационная связность, достаточно высокая структурированность и качество информационного наполнения. К таким ресурсам можно отнести, например, тематические порталы, в том числе учебного назначения. При наличии развитой онтологической компоненты можно говорить уже не просто про поиск по классическим схемам, а про автоматизированный экспертный подбор наиболее адекватных информационных ресурсов, и такие системы должны приобретать экспертно-консультационный характер [4].

В работах [4, 5, 6 и др.] развивается подход на основе построения и анализа формальной модели информационного наполнения тематического веб-портала в виде графа “онтология-документ”. В рамках этой модели вводится определенная система параметризованных математических соотношений, описывающих взаимосвязи между различными тематическими узлами и соответствующими им документами. В то же время остается открытым вопрос о выборе параметров этих соотношений. Очевидно, что априорный подбор этих параметров не представляется возможным, и они должны выбираться и корректироваться уже в рабочем режиме, экспериментальным путем. Возникает необходимость в создании инструментальных средств, обеспечивающих проведение подобного эксперимента и анализ его результатов [4].

В этом контексте как один из возможных подходов, направленных на проектирование и создание инструментальных средств, обеспечивающих качественную настройку системы информационного поиска, рассматривается концепция модельной поддержки информационного поиска [7]. Речь идет о создании некоторого банка моделей, описывающих различные аспекты поведения системы, а также связанных с этими моделями интеллектуальных программных средств. Сами по себе элементы системы модельной поддержки тоже могут быть связаны друг с другом и, соответственно, образовывать те или иные структуры (например, в виде деревьев или графов). Настоящая работа направлена на рассмотрение возможных компонентов модельной поддержки информационного поиска и связей между ними.

Основное содержание работы

Как базовая рассматривается модель информационного наполнения тематического портала в виде графа “онтология-документ”, построенная на основе формальных моделей онтологий [4-6 и др.]. Эта модель в наиболее общих чертах описывается как тройка $M=<W^*,D,L>$, где W - онтология предметной области, W^*-расширенная онтология, наполнение онтологии W конкретными экземплярами классов (фактически – база знаний), D - множество документов; L – множество связей между W^* и D. Собственно онтология описывается как тройка $<Q, R, F>$, где Q – множество классов, которые соответствуют понятиям предметной области, R – множество связей между ними, а F - множество функций интерпретации. Соответственно, расширенная онтология описывается как тройка $<Q^*, R^*, F^*>$, где Q^* - множество классов вместе с их экземплярами, R^*- множество связей между этими элементами, а F^* – множество

функций интерпретации, определенных в простейшем случае на элементах из Q^*, R^* та Q^* x R^* x F^*. Тогда элементы D могут быть значениями функций из F^*. Такая формализация описывает граф "онтология-документ", узлы которого соответствуют понятиям предметной области и информационным ресурсам, а дуги – связям между ними, причем эти связи могут быть разных типов. Таким образом, если w является элементом расширенной онтологии, а d – артефактом информационной системы, то функции интерпретации f и соответствующие весовые коэффициенты могут формироваться на основе этих категорий сущностей.

Далее, может быть осуществлен переход к модели "онтология-артефакт-пользователь-проект", в которой меры важностей связей зависят от характеристик и целей посетителей. Альтернативный взгляд на проблему может заключаться в построении многокомпонентной онтологической системы, отдельные компоненты которой соответствуют отдельным категориям сущностей [8]. Все это позволяет в общих чертах охарактеризовать наиболее важные компоненты системы модельной поддержки информационного поиска, в частности:

1. Базовая модель в виде графа "онтология-документ", описанная выше.
2. Параметризованные модели, описывающие собственно меры близости между узлами онтологии предметной области и документами. Для построения таких мер близости можно использовать ряд известных подходов на основе булевой и взвешенной векторно-матричной модели, теоретико-множественного анализа связанных элементов [1-3 и др.].
3. Параметрические модели, задающие комбинированные меры релевантности. Обозначим через $r_q(w,d)$, $q \in Q$, $w \in W$, $d \in D$ меру релевантности документа d понятию w по связи q. Здесь W – множество понятий предметной области, D – множество артефактов информационной системы, Q – заданное множество возможных типов связей. Тогда естественно рассматривать некоторую комбинированную меру релевантности документа d поняттю w, усредненную по всем связям с учетом их весовых коэффициентов [4]:

$$R(w,d) = \sum_{q \in Q} \alpha_q r_q(w,d) \qquad (1)$$

где α_q – вес (содержательно - мера важности) q-го типа связей.

4. Модели, определяющие собственно процесс принятия решений относительно динамического формирования навигационного графа, задающего возможные переходы между узлами веб-сайта.
5. Модели, характеризующие посетителей веб-сайта и позволяющие построить профили и выделить наиболее типичные группы пользователей.
6. Модели, характеризующие возможные цели посетителей. В частности, в [1] описаны такие базовые типы запросов, как информационные, навигационные и транзакционные. Эти модели позволяют относить запросы к той или иной группе и на этой основе принимать более обоснованные решения о выборе материалов, наиболее релевантных запросу.
7. Модели, описывающие поведение пользователей и их навигацию по сайту. Этот вопрос в общих чертах обсуждался в [6]. В частности, в [6] утверждается, что рассмотрение и анализ некоторого порождающего навигационного процесса позволяет

получать семейство соотношений, аналогичных соотношениям классического алгоритма PageRank, но уже с учетом семантических связей на графе "онтология-документ".

8. Модели, описывающие взаимное влияние между узлами. В основе подобных моделей лежит интуитивное соображение о том, что мера важности узла может зависеть от мер важности связанных с ним узлов. В частности, этой основе в [5] описывается методика динамического перераспределения мер важности узлов на основе организации волнового процесса распространения активации.

9. Модели, описывающие собственно процесс обучения и настройки системы. Для подбора коэффициентов соотношения (1), в частности, могут применяться генетические алгоритмы [9 и др.], хорошо зарекомендовавшие себя при решении многих переборных задач. Некоторые подходы к применению генетических алгоритмов к задаче оптимизации информационного поиска в общих чертах описаны в [10].

10. Следует обратить внимание на модели и методы Data Mining, то есть интеллектуального анализа данных; поиска закономерностей, которые объясняют имеющиеся данные; добычи знаний из сырой информации [11]. Для задачи информационного поиска на тематическом портале особое значение имеют методы Web Usage Mining [12], выделяемого как самостоятельное направление и связанного с анализом посещаемости веб-ресурсов и выявления закономерностей, которые объясняют поведение посетителей.

В частности, на основе методик Data Mining можно ставить вопрос о выборе оптимальных параметров соотношения (1). Действительно, можно предполагать, что пользователь выбирает ссылки, которые он считает наиболее перспективными, и тогда основой для выбора параметров (1) становится история фактически осуществленных переходов по ссылкам.

В рамках описываемого онтологически-ориентированного подхода можно рассматривать такие постановки задач Web Usage Mining:

- множество посетителей разбивается на кластеры или по своим профилям, или по истории навигации; для каждой группы определяются наиболее приоритетные типы связей между узлами графа "онтология-документ", и на этой основе расставляются персональные весовые коэффициенты, которые зависят от характеристик посетителей;
- на основе анализа истории переходов между узлами графа "онтология-документ" оценивается вероятность того, что находясь в узле q с определенным значением характеристики a, посетитель перейдет по ссылке, которая соответствует типу связей r;
- оптимизация структуры навигационного графа с целью сокращения последовательности переходов, которые должен осуществить пользователь, чтоб достичь цели;
- эффективный подбор контекстной рекламы, связанной с ресурсами с наивысшей оценкой релевантности - то есть с теми, которые могли бы с наивысшей вероятностью заинтересовать посетителя, который в данный момент находится в заданном узле графа "онтология-документ".

Построение системы модельной поддержки информационного поиска, кроме собственно набора моделей разных типов и связанных с ними процедур, должно

предусматривать организацию связей между ними. В частности, следует предусмотреть:

- объединение отдельных моделей в сеть, на основании которой можно осуществлять целенаправленные переходы между ними с целью поиска наиболее подходящих из них;
- механизмы автоматического запуска тех или иных программ, связанных с моделями как узлами модельной сети.

Выводы

В работе в общих чертах описываются возможности, связанные с организацией модельной поддержки онтологически-ориентированного поиска на тематическом портале с целью повышения его эффективности, точности и релевантности. Базовая модель информационного наполнения портала на основе графа “онтология-документ” должна быть погружена в сеть моделей. Уточнения и дальнейшие формализации рассматриваемого подхода являются предметом дальнейших исследований.

Литература

1. Маннинг К.Д., Рагхаван П., Шютце Х. *Введение в информационный поиск.* – М.: ООО «И.Д. Вильямс», 2011. – 528 с.
2. Гаврилова Т.А., Хорошевский В.Ф. *Базы знаний интеллектуальных систем.* – СПб: Питер, 2000. – 384 с.
3. Ландэ Д.В. *Поиск знаний в Интернет*. – М.: Изд. дом «Вильямс», 2005. – 272 с.
4. Олецький О.В. Організація онтологічно-орієнтованих засобів автоматизованого добору інформаційних ресурсів на тематичному порталі. *Наукові записки НаУКМА. Т.99. Комп’ютерні науки.* – К., 2009. – С. 66-69.
5. Олецький О.В. Онтологічно-орієнтований інформаційний пошук на основі хвильового процесу поширення активації. *Наукові записки НаУКМА. Т.86. Комп’ютерні науки.* К., 2008.,50-52.
6. Олецький О.В. До проблеми моделювання потоку відвідувань на онтологічно-орієнтованому тематичному порталі. *//Моделювання та інформаційні технології. Збірник наукових праць. Спеціальний випуск.* Т.2.- К.,2010, 321-326.
7. Верлань А. Ф., Дячук А. А., Сагатов М. В. *Модельная поддержка автоматизированного проектирования сложных технических систем.* М., 2008.
8. Олецький О.В. Побудова багатокомпонентної онтологічної системи для автоматизованого експертного підбору навчальних матеріалів. *Теоретичні та прикладні аспекти побудови програмних систем.* Матеріали міжн. конференції TAAPSD'2009. Київ, 83-88.
9. Рутковская Д., Пилиньский М., Рутковский Л. *Нейронные сети, генетические алгоритмы и нечеткая логика*. – М.: Горячая линия - Телеком, 2004. – 452 с.
10. Олецький О.В. Принципи застосування генетичних алгоритмів до задачі онтологічного інформаційного пошуку. *Наукові записки НаУКМА. Т.112. Комп’ютерні науки.* К., 2010, 49-54.
11. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. *Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP*. – СПб: БХВ-Петербург, 2007. – 384 с.
12. Гончаров М. *Web Mining – добыча знаний из World Wide Web.* http://www.spellabs.ru.

Quasi-Smart Construction Synchronous Machine Working with Extensive Cycling Load

P. G. Ali-Zade[1], R. N. Tuncay[1], S. B. Ozturk, O. C. Kivanc

[1] Okan University, Istanbul, Turkey

Abstract. The paper deals with: 1) The controlled exciting systems to keep cosφ ≈1 of the special brushless synchronous motor (SBSM) under cycling load. They are firstly unmanned the deep-well oil pumps drives, wind renewable energy synchronous generators, etc. The machines usually work in harsh climate conditions such as extreme northern cold or deserts hot outside ambient temperatures (rarely up to ±60°C), where traditional electronic controlled exiting systems are not reliable (except diodes). 2) Peculiar approach during project work and construction of the special smart brushless synchronous machine (SBSM); it includes their designs and industrial applications with the simple diode bridge rectifier exciting system with constant current to get self-stabilizing effect of power factor (cosφ ≈1) under wide range of cycling torque; it was made by only special modifications of the motor to put it in smart state for oil field sucker-rod pumps drives.

Keywords. Energy saving, induction motor replacement, self-stabilizing power factor, sucker-rod pump, brushless synchronous motor, exciting system.

I. Introduction

This paper outlines the designs and applications of special brushless synchronous machine (SBSM) mainly for oilfield sucker-rod pumps (SRP), or wind and ocean wave renewable energy shown in Figs. 1 (a), (b) and (c), which suppose to keep average cosφ ≈ 1. The studies cover and based on the following aspects and issues:

1) The oil field equipment are usually unmanned and operating in severe and harsh environmental conditions such as extreme cold or hot outside ambient temperature (rarely up to ±60C);
2) The technical and economical reasons for the replacement of distant SRP induction motors (IMs) by the SBSMs, while diminishing harmful effects on the environment (as reliable as IM), are to save energy, improve stability, reliability, and reducing of the oversized motors use
3) The special approach during the SBSM design work to get smart self-stabilizing cosφ ≈ 1 effect under a wide range of cycling torque
4) The practical implementations of the proposed designs in several oilfield experiments and computer-assisted electro-dynamic lab model studies: IM and SBSM dynamic stability analyses under their supply voltage wide fluctuations and extensive cyclic moment on the shafts

1) Equipment often has to work in inaccessible places far away from their technical service centers (arctic locations, tundra, hot or cold deserts, sea jack-up derricks, real or artificial islands in swamps or in lakes, etc). In most cases there are high requirements for reliability and survivability of the machinery, since the equipment often may be serviced only once (rarely twice) a year by mobile team of technicians (using a helicopter, ship or all-terrain track-type vehicle, etc).
2) Any unexpected temporal equipment outage even for 10-20 minutes, due to any reason (mechanical, electrical, electronic, etc) can lead to an array of problems for extended periods, especially during winter - frozen well liquid in output pipes, lubricating oil in SRP reducers, etc.
3) In the so-called reach of sand wells scenario during any stoppage can be dramatic: sand settles gradually down by gravity and blocks SRP subterranean reverse valves. It makes automatic "self-start" of the pump more difficult and longer or even impossible, which may lead to a forced outage up to the next service team visit.
4) One of the main requirements of any gas or oil field installation: it must not have in or around every well any sparking elements, which could provoke fire or explosion. Drives must have brushless motors, fire or explosion-proof switches, control systems, etc.

It is shown in the paper that the special brushless synchronous motors (SBSMs) help to cut energy losses in oil field electric network, to reduce oversize motor utilizations, to increase motors group stability, reliability and even to reduce negative ecological impacts.

The calculated SBSM production cost is 130-150% of the IM. It is estimated that the SBSM additional manufacturing cost can payback in about 4–6 months.

Some quantitative and economical estimates, which have been achieved on the basis of several field experiments and by computer-assisted mechatronics dynamic modeling, are presented in the papers [1, 2, 3, 4, 5, 6, 7 and 8]. It also includes IM and SBSM dynamic stability analyses and comparison in conditions of supply voltage fluctuations and heavy cyclic moment on their shafts.

II. Discussion

Generally, oil well SRPs are widely dispersed around their main transformer that is usually tapped up to maximum. The ordinary sucker-rod pump (Fig. 1a) motor works under the extensive specific cycling moment on its shaft (Fig. 1b) and, in addition to, distant SRP under considerable deep voltage drops and fluctuation (Fig. 2) of up to 20% of V_{NOM} = 380V and usually reach about ΔV = 76–80V, but seldom can be even greater. As a result, there are big substantial power losses in the field electrical network, and poor group stability of the motors, etc. Thus, the oil industry needs highly reliable and survivable motors for its fields, which could be serviced, as mentioned above, practically only once a year by a special mobile service team.

Induction motors (and compensating capacitors for power factor improvement) are widely used as electric drives for the sucker-rod pumps (SRPs) as well as in Azerbaijan oil production industry, and in other former USSR republics, particularly in Tataria and Siberia, oil fields, etc. Approximately 70–80% of all acting wells are equipped with SRP (but produce only 30–40% of oil). SRP will be a known "destiny" for many dozens of years in every oil well when its oil production becomes poorer.

The analysis of SRP oil field electric power consumption over the last 20–30 years has shown that the consumption per ton of oil produced is increasing [3, 6, 9, 10]. This growth is attributed by effects shown on Figs. 1(a),1(b), and 2.

These voltage fluctuations are due not only to the character of the load on the shaft, but also due to the neighboring IM cycling loads. The sharp voltage drops are following the most of self-starts of one or a group of electric motors, distant short-circuits in the supplying network, and the different regime commutations in the high voltage power system etc. All these have a significant effect on the electrical and economical characteristics of the IM used.

Normally the IM torque Fig. 1(b) varies from values close to the no-load torque of about $M_{MIN} \approx 0.1–0.2$ (for poor balanced SRP comes down to negative $M_{MIN} \approx – 0.3$) up to overloads $M_{MAX} \approx 1.4–1.6$ of the rated (nominal) torque M_{NOM} with frequencies between 3–15 cycles per minute ($f \approx 0.05–0.25$Hz). The cyclic loads, phases and frequencies of SRP differ from each other and cause the voltage across the IM terminals to fluctuate in a very complex manner (Fig 2). This further impairs the performance of oil field electric power network, poor as it is. Voltage fluctuations reach 15%, and short-term voltage drops on supply line may be up to 20% with duration of around 0.4 s (rare even more). The harsh environment and mentioned above combined effects often lead to frequent self-overturning and stoppage of induction motors.

In order to increase the reliability, oil fields frequently have to use oversized motors (Fig. 3) and exceed the next two or even three standard higher sized motors, which obviously results in low cosφ, extra material and electric energy losses up to 30-40 % of the total electric power consumed by the SRP oil field.

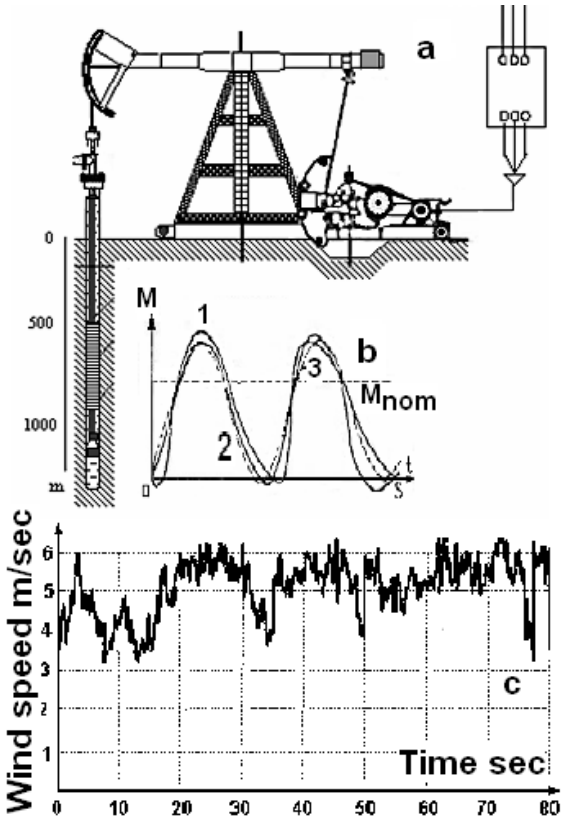


Fig. 1. (a) Sucker-rod pump. (b) Moment on the SRP shaft: 1- field oscillogram, 2 - calculated, 3 - sinusoid, (c) real extensive variable wind speed via time oscillogram.

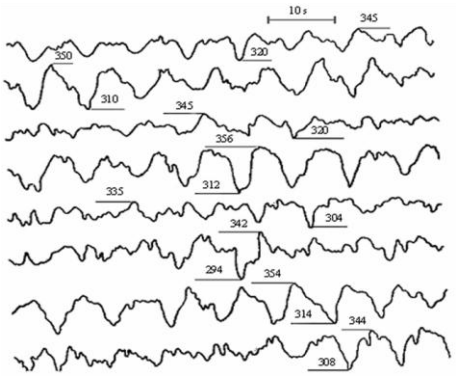


Fig. 2. The actual voltage fluctuation of oil field sucker-rod pumps.

About 200 distant SRP IMs were tested (10 IM –7 kW, 43 IM –10 kW, 35 IM –14 kW, 68 IM –20 kW, 33 IM –28 kW). Results can be seen in Fig. 3, vira (maximum) currents

of about only 10% motors normal and are higher than their rated INOM currents. They all are supposed to be approximately 1.4–1.6 times higher than their INOM for warm regions of the countries. Thus, about 90% of IMs of northern oil fields are oversized in the Summer, i.e. their maximum torque on the shaft is two - three times less than the nominal one of the motor and low cosφ ≈ 0.5–0.55. However, during cold time, maximum torque on the shaft increases due to freezing of the SRP reductor lubricating oil, which partially can be justification of the oversized motors use.

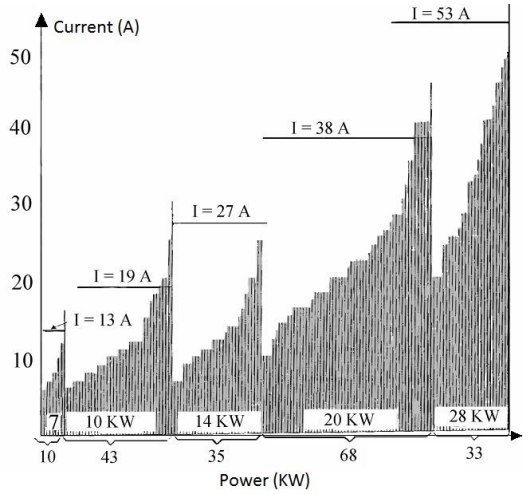


Fig. 3. Oil field IM maximum (vira) current loading rates.

III. Smart Self-stabılızıng cosφ ≈ 1Brushless Synchronous Motor for Extensıve cyclıng Load

Several types of special brushless motors (SBSM) were elaborated for oil industry and applied [2, 4, 5, 6, 7 and 8] in the Azerbaijan State Oil Company "AZNEFT" oil fields. Feasibility studies of the SBSM used as the SRP electric drive were conducted twice in support of industrial production of the SBSM. The rated power of the motors varies from 3.0 to 50 kW. The most widespread is 20 kW for Azerbaijan (50 kW for some Siberian and Tataria oil fields).

It could be a bit strange from control point of view, but the main proposal of this paper is how to avoid any controlled electronic exciting system of usually controlled synchronous motor SM, but in addition to get stabilizing cosφ≈1 effect by only special construction of the motor.

Contrary to popular control science theory, the second main thesis of this paper is to avoid any exciting current thyristor-transistor control system (except simple diodes) of the usually controlled synchronous motor. At the same time, SM as electric drive must usually keep cosφ ≈1 during wide scale moment deviation, in other words, – to show "power factor cosφ ≈1 self-stabilizing effect". That works to the effect of minimizing power losses in the

supply line (network). This advantage can be obtained with a small constructive change to the SM stator. The constant exciting current of the simple diode tree or six pulse rectifiers are much more reliable than a thyristor-transistor controlled one. It also raises SM total reliability [9]. SBSM rotor consists of two monolith pole systems on its shaft with the special claw-type trapezoidal heads on them. A squirrel cage helps to keep the poles together and to improve the starting characteristics of the motor.

The shaft is made of non-magnetic steel (or there is non- magnetic steel part cut-weld into the shaft). Two exciting coils are placed over bearings inside the back and front covers of the motor. There is no any transformer in the power supply of exciting system. Power comes directly from a 0.4 kV main to rectifier. Therefore, there is also a big flat copper ring disk washer on or under each of the coils to reduce the induced electromotive force (emf) jumps in the coils during start of the SBSM (without the washers emf jumps up to 800 - 1200V).

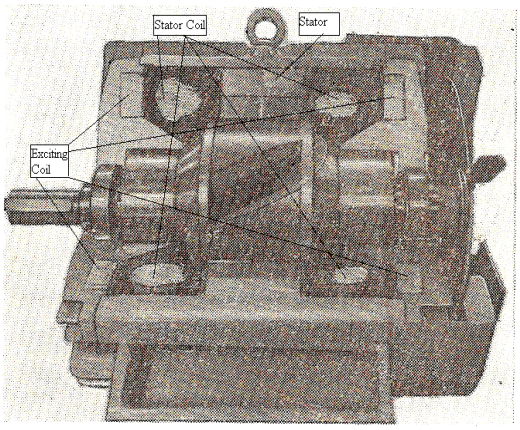


Fig. 4. Brushless synchronous motor (generator).

The distances between the exciting coils and the stator are bigger than in a normal SM. There is some additional flux leakage that a bit raises IEX nominal.

Such construction of the SBSM provides cosφ ≈1 for a wide range of power (torque) deviations. The SBSM can start at any position of the pump system and the motor enters into synchronism within 3–5 s, when the SRP cycle load moment comes down, and it is less than the SBSM synchronization moment (for normal SBSM MSYN ≈ 0.25–0.4).

The theoretical and experimental analyses of the SBSM cos φ characteristics variations due to small alterations of its stator length show [2, 3] that the shorter the stator of the SBSM, the flatter around cosφ = 1 its cosφ versus power characteristic will be (Fig. 5). The corresponding rises of stator magnetic power losses of the SBSM are negligible small in respect to the above mentioned power losses in the oil field networks.

TABLE I
VALUES OF PHASE VOLTAGE, PHASE CURRENT AND POWER

Parameters	IM MA-144-2/4	SBSM CDBPK 81-(IEX ≈2A)	Shaft position SRP N0 2884
Phase Voltage (V)	215 208 214 210	228 225 222 225	00 900 1800 2700
Phase Current (A)	21.0 40.0 25.2 48.5	7.4 29.2 8.7 38.2	00 900 1800 2700
Power (kW)	-2.8 18.2 -6.0 22.1	-3.6 19.8 -4.2 23.4	00 900 1800 2700

Thus, seeking the cosφ ≈ 1 self-stabilizing effect has been found to meet the requirements: while allowing the stator of the SBSM to be ≈5% shorter, with no other special distinctions from an IM or SM stator.

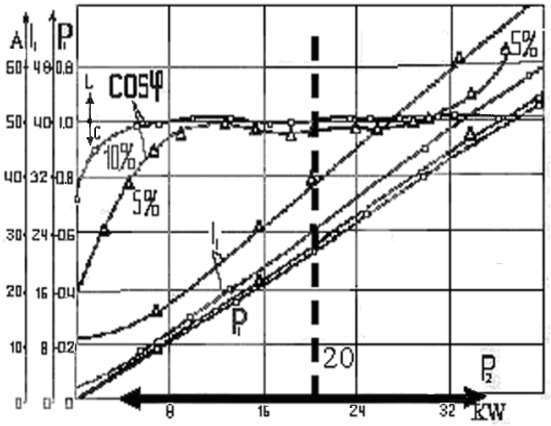


Fig. 5. Special brushless synchronous motor main characteristics.

As far as wrongly balanced SRP (with negative MMIN) is concerned, the SBSM generator regime characteristics are almost the same and it keeps cosφ ≈ 1 around nominal power, as well. Therefore, the SBSM characteristics make them suitable for use as a wind farm (subsea stream) generator, for unmanned, distant automatic meteorology and oil (gas) pipes anticorrosion stations among other applications, too.

The rated power of the last model designed SBSM varies from 3.0–50 kW. The most widespread power is 20 kW. Two IM of 28 kW, distant from main transformer, were

replaced by the last designed 20 kW SBSM model on the SRP of "AZNEFT" oil wells №2884 (depth 645 M, 8 cycles a minute, slightly misbalanced) and the second - on №2033 (depth 900 m, 8 cycles a minute). The third IM of 3 kW was placed into the electrodynamics lab model for the computer-assisted stability analyses.

The survivability test of the SBSM, when it loses exciting current ($I_{EX} = 0$), one of the SBSM was left without a rectifier bridge (exciting coil was short circuited) and started to work in the asynchronous regime from spring to autumn. As a result, it has been working in asynchronous regime successfully for more than 6 months even during all summer in Azerbaijan.

Field Results: The cycling current (Table 1) came down from $I_{MIN} \approx 21–25A$ and $I_{MAX} \approx 40–48.5A$ for IM (average I_{AV} IM $\approx 44.24A$) to $I_{MIN} \approx 7–9A$ and $I_{MAX} \approx 29–38A$ for SBSM (average I_{AV} SM $\approx 33.5A$), respectively. Power losses in the motors supply line came down 1.7–2 times mainly during maximum current of a cycle (9.4 times during minimum current).

The calculated commercial production cost for 10,000 SBSM a year (for 16–50kW) is 30–50% higher than IM of the same power. After all lab tests and several years industrial exploitation of different SBSM models the feasibility study was specified and corrected. The payback (compensation) time of the SBSM was determined to be between 4–6 month due to the solution of oversize motor problems and the reduction of energy losses (common rise of network voltage, better motors group stability, reliability, conditions of the neighboring motors and ecology improvement factors were even not taken into account).

IV. Conclusıon

The production cost of the SBSM (for 16–50kW) is 30–50% higher than that of IM with the same power rating. The laboratory and field experiments with unusual types of SBSM exciting systems for different climatic zones have shown their reliability and vitality. In the studies, it was determined that the payback (compensation) time for the increased SBSM production cost is between 4 to 6 months taking into consideration only the solution of oversized IM motor problem and a reduction of energy losses. Common rise of main voltage, better motor group stability, reliability, as well as conditions of the neighboring motors and ecology improvement factors were not taken into account due to difficulty in economical calculations, but these factors certainly represent the significant additional benefits of the SBSM implementation.

The paper presents SBSM construction system of oil field sucker-rod pump drive which keeps $\cos\varphi \approx 1$, where traditional transistor-thyristor or IGBT controlled exiting systems are not reliable (except diodes). Therefore, the SBSM characteristics make them suitable for use as electric drive in oil industry, but can be also used as a windmill farm (subsea stream mill) generator, for unmanned, distant automatic meteorology and oil (gas) pipes anticorrosion stations, for group wind farms of distant oil fields among other applications.

V. Acknowledgments

The authors wish to thank the individuals who supported and provided real practical help in workshops, on oil fields and in the labs to study and develop several pilot projects of

SBSM, to OKAN and Istanbul Technical University colleagues and to all, who have shared their comments, corrections and suggestions during several meetings and conferences.

VI. References

[1] V.M. Kutzelov and L.A. Kovaluk, "Electromagnetic calculations of brushless synschronous motor with external core," *Znanie Publ. H.*, Riga,Latvia, 1973.

[2] K.N. Kulizade, P.G. Ali-Zade and I.E. Khajkin, "Synchronous electric drives with pulsing load," *Energy Publ. H.*, 1973.

[3] P.G. Ali-Zade et al., U.S.S.R. Patents: SU 392514 (1973b), 416707 (1973c), 560655 (1975), 698013 (1979), 955115 (1982).

[4] P.G. Ali-Zade, A.S. Kuliev and V.A. Fedorchuk,"Studies into the electric drive dynamics of an oil production mechanism by the structural modeling on a personal computer, " *Engineering Simulation Journal*, vol. 10, pp.1077-1085, USA, 1993.

[5] P.G. Ali-Zade and O. Ustun, "Oil industry energy conservation and management study by hybrid computer-electro dynamic model and oil extraction electric drives, " presented at the Int. Conf. Proceedings on Energy, Kuwait, 1998.

[6] P.G. Ali-Zade et al., "Multi-functional computer-assisted electro dynamic model for industrial mechatronics simulation, " presented at the Int. Conf. Proceedings Mechatronics, Atlanta, USA, 2000.

[7] P.G. Ali-Zade and E. Imal, "Approach to construction of a synchronous motor with constant exciting current to get self-stabilizing cosφ = 1for wide range cycling torque, " presented at the Int. Conf. Control, Cambridge, U.K., 2000.

[8] P.G. Ali-Zade et al., "Motor's expense repaid in savings," I*EEE Industrial Applications Journal*, pp. 21-28, 2006.

The Search for Adaptive Methods Filtering Out High-Noises and Their Application on Oil Complexes' Control Telemetry Systems

O. Ustun[1], M. Yilmaz[2], P. Ali Zada[3], R. N. Tuncay[3]

[1] Electrical Engineering Dept., Istanbul Technical University, 34469, Maslak, Istanbul, Turkey

oustun@itu.edu.tr,

[2] Electrical and Computer Eng. Dept., University of Illinois at Urbana-Champaign, IL, USA

myilmaz@illinois.edu

[3] OKAN University, Akfirat Campus, 34959, Istanbul, Turkey

parviz.alizade@okan.edu.tr, nejat.tuncay@okan.edu.tr

Abstract. This article's domain is the application of well-known modern applied mathematical methods in a control structure of oil industry engineering systems. Different types of Electrical Submersible Pumps (ESP) are running on worldwide oil fields. The ESP telemetry system allows for the obtaining of information in the neighborhood of different heavy electromagnetic noises such as: random, pulsing and harmonic. The telemetry system needs to receive accurate information on the pump unit's intake pressure, temperature and most importantly for the submersible motor, the stator cooling oil insulation resistance, for the successful exploitation of the ESP. The change from analog to digital telemetry techniques is moving quickly in almost all technical fields. But increasing disturbance levels with the corrupted analog telemetry signal are resulting in increasing noise levels; however, often it is still 'audible' or the control system is still reliable. Though, beyond a certain disturbance level, the so-called 'digital cliff,' the digital telemetry signal and control may stop abruptly. Here in the paper, an analog signal processing implementation was researched for the detection of the most efficient adaptive noise-cancelling filters among dozens of recognized ones for oil industry ESP telemetry control systems of under severely noisy conditions. Unfortunately, not all of the noise-cancelling methods showed good results in the ESP digital model of the mentioned case of different types of heavy electromagnetic noises. From ten applied adaptive filter algorithms, only three have shown successfully good results in the early prediction of the ESP motor real insulation disruption (like *Sign-error, Sign-data* and *Sign-sign* filters). The best among the ten analyzed adaptive filter algorithms was recognized to be, The Normalized LMS FIR filter algorithm - *adaptfilt.nlms*.

Keywords: Signal, noise, adaptive methods, oil industry, submersible pump, communication-telemetry channels.

1. INTRODUCTION

More than a thousand switchboards of Electro-submersible Pump (ESP) under different trademarks are running in the oil fields of modern Russia and the CIS (former USSR), representing a wide spectrum of varied equipment which are working on the problems in oil production and its optimization. Not only are there simple devices of ESP motor control, but also there are complicated electronic complexes for installing an oil well borehole for operational duty as well. The main problem is to preset supporting parameters during oil production at several well borehole clusters simultaneously. At the same time, the ESP operating information can be read out on any company computer, on any data transfer and storage micro-unit. Moreover, it can also be transferred to a telemetric system by a special port or radio channel to a dispatcher control board for further analyses of the pump unit and database support. The ideology of the equipment consists in an integrated approach towards the automation and optimization of the oil production and an intention to produce oil production equipment to compare the well data with other samples in the world. The submersible telemetry system usually allows for the obtaining of information on the pump unit's intake pressure, temperature and most importantly for a submersible motor stator coil, its insulation resistance, for the successful exploitation of the oil complexes in the neighborhood of different types of heavy electromagnetic noises such as: random, pulsing, harmonic, etc. [1, 2 and 3].

The change from analog to digital communication techniques is moving quickly in almost all technical fields. The emission limits for radiated and conducted disturbances, as prescribed by international standards, are based on their possible impact on analog telemetry. When the existing analog telemetry technologies are substituted with digital technologies, interference could occur even though all equipment complies.

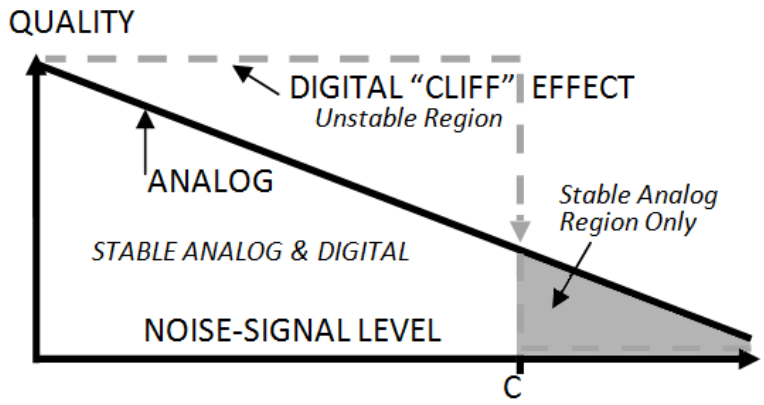


Fig. 1. Digital cliff, unlike analog transmissions which experience gradual degradation, can experience sudden degradation.

Increasing disturbance levels with corrupted analog telemetry result in an increasing noise level, but it's often still 'audible' or the control is still reliable. Digital telemetry technology is much better in suppressing disturbances up to a certain level. With increasing disturbance levels, the analog signal will remain low, but audible. However, beyond a certain disturbance level, the digital telemetry and control stops abruptly. This, the so called 'digital cliff' at point *C* (shown in Fig. 1), makes it more complicated with digital telemetry to know the 'headroom,' (or the disturbance critical level), before the digital telemetry and control unexpectedly stops. The analog signal after point *C* will be lower, but still audible, and the control system is still reliable.

Here, an analog signal processing implementation is studding for the detection of the most efficient adaptive noise-cancelling filters among dozens of well known ones for telemetry control of oil industry power complexes under severely noisy conditions. A useful approach to this filter-optimization problem is to minimize the mean square value of the error that is defined as the difference between some desired signal and the filter's actual output.

There are many noise cancellation methods and applications in industrial, civil, military, power systems telemetry and the control equipment's apparatus. But the success of these noise- cancelling methods and filters depend mostly on the noise factor (signal/noise ratio) and also on the control signal character under consideration: close to random, exponential, voice, sinusoidal, etc.

Most of the publications in the field of noise cancellation methods and their applications deal with rather big signal/noise ratio (that is – noise ‹‹ signal) and show good achievements for cell phones, radio/TV technique, tape recorders, apparatus for people with hearing loss and concert hall equipments.

There are some cases when the signal/noise ratio is around 1 (noise ≈ signal): command-and-control telemetry systems between centers and operators of winches, tractors, lorries, textile factory workshops, compressors workshops, metro, railways, controlled AC/DC drivers, as in the submersible pump telemetry case and so on. Not all of the noise-cancelling methods show good results, especially in real-time noise filtering, and need a special study to find the best filtering method for each engine's particular noise situation.

The special cases are some severe ones which deal with the signal/noise ratios ≈ 0.2÷0.1 (noise ›≈ 5÷10 higher than useful signals): military command-and-control telemetry between centers, jet pilots and jets service teams of aircraft carriers battleships, in different metallurgical and especially arc furnaces telemetry between dispatcher and operators team, etc. There is an understandable shortage and vacuum of any information and publications not only in military fields, but also in the modern technological companies as well.

A similar situation of low signal/noise ratio is presented in the field of the power systems, powerful controlled electrical machines and high voltage power electronics control-feedback signals. Any wrong interpretation of the signals due to high noise can bring unexpected accidents or malfunctions of large power systems or heavy duty electromechanical installations.

Due to the mentioned 'digital cliff,' the manufacturers of powerfully controlled electrical machines unfortunately still have to sometimes work with analog signals and equipment. This paper deals with the oil industry ESP motor, in particular. with its control which is working close to, or in the neighborhood of, different heavy 'jam' of electromagnetic noises such as: random, pulsing, harmonic and so on, that overwhelm (engulf) the useful signals.

As will be shown, for the ESP case, only a few of the *MATLAB* noise-cancelling methods - *Adaptive Filtering Methods* - present good real-time noise filtering results for the mentioned ESP severe noise cases.

Every other apparatus case needs a special study to find the best filtering method for the particular equipment [4, 5, 6].

2. THE OIL INDUSTRY ELECTRICAL SUBMERSIBLE PUMPS

Electrical Submersible Pump (ESP) induction motors are applied as a drive for electrical centrifugal oil, deep-well pumps. They are usually put on the market in diameter/dimensional groups of 103, 117, 130 and 180 mm, and power ranges from 12 to 500 kW, and even more. There are more than 60 different modern types of ESP of various capacities, which allow for picking up the optimal motor-pump to get the maximum possible efficiency for the particular oil field [7].

The techniques manufacturers have to keep are of the highest quality and performance reliability of the ESP motors because of:

1. Very expensive equipment often has to work in inaccessible places or faraway from their technical service centers (the Arctic tundra locations, hot or cold deserts, sea jack-up platforms, real or artificial islands on the surface of swamps or lakes, etc.). In some cases, there are high requirements for dependability and survivability of the machines, since the equipment often may be serviced only once a year by a mobile team of technicians (using helicopters, ships or all-terrain track-type vehicles, etc.).

2. Any unexpected temporary equipment outage even for 10-20 minutes for any reason (mechanical, electrical, electronic, etc.), can lead to an array of problems for extended periods, especially during the winter season, such as frozen well liquid in output pipes, lubricating oil, etc.

3. In the so-called reach of sand (or clay) wells, the scenario can be even worse as any stoppage can be dramatic and sand may gradually settle down by gravity and block the ESP subterranean reverse valve at the bottom of the well. This makes the automatic ‘self-start’ of the pump more difficult and longer, or even impossible, which in the winter may lead to a forced outage up until the next service team’s visit (which could be the next spring or even summer in Siberia, Canada, Alaska, etc.). Moreover, it requires an expensive and complicated lift of the pump exploitation column and the cleaning of the ESP pipes and reverse valve from the sand and other debris.

To raise working survivability and to keep the highest quality, some ESP manufactures apply new techniques and measures, for example:

1. A stator made with the closed groove that raises the cleanliness of the motor’s internal room of the engine and allows for the successful application of the winding grooves’ firm insulation in the tube form.

2. The electric motor rotor has original bearings, having the mechanical fixings from any cranking.

3. The application of special, modern electro-technical materials allows for maintaining ESP motors at temperature strata liquids up to 120°C and with superheat-resistant materials up to 160°C.

4. 100% of ESP’s have to be disassembled for all elements to be checked, and, after all of these have been checked, the ESP parts are thoroughly re-assembled and the ESP must

be carefully tested in conditions similar to the real ones, including heating the ESP up to working temperatures, as would be in a well.

After 1-hour of the above-mentioned check up, the motor winding isolation's normal resistance should not be less than 10 MOhm. The control of the normally working motor resistance of the system 'transformer-cable-ESP' isolation must not be less than ~350 kOhm. This isolation resistance usually decays very slowly over several months (as a rule - exponentially) from ~350 kOm to ~30 kOhm – but not less. The decaying process can take several months, even years, but the resistance of less than ~30 kOhm is potentially risky and can be a provocateur of any heavy internal short circuit in the stator winding and destroying some of its section. This means that it will be an expensive and complicated lift of the exploited column, removal of the burnt section of the motor, cleaning of the burnt section on the test bench and then the re-wiring of it. Then again**,** after disassembling each component, the quality of each must be supervised and the electrical motor must be tested at the workshop station, etc. Thus the ESP motor is the most sensitive and most expensive element of the ESP and control of its working motor insulation resistance must be very strong and effective [8, 9].

2.1. Structural Cart of the ESP Unit Equipment Complex

The electrical submersible oil pump (ESP) is usually fed from a 400 V network and consumes currents in-between 160 – 1200 A. The ESP consists of the following main parts: switchboard, step-up transformer, variable speed driver (VSD) convertor with built-in electronic active filter (FSA) and electrical submersible oil pump (ESP).

ESP has a very complicated telemetry system. Among other ESP parameters, only the special ESP electric motor telemetry controls the following motors main parameters**:** R–insulation (resistance) and t_o–temperature (^oC) of the ESP motor cooling oil**,** the ESP motor axial and radial vibrations and also some oil pump technological parameters like the formation liquid t_p–temperature (^oC) on the pump discharge, P–pressure on the pump discharge and the formation of the pumping liquid flow, etc.

As it was mentioned, the ESP electric motors are very sensitive to their regime parameters. It is the cooling oil-filled motor (seldom distilled water), separated by a strong gasket from the pumping liquid, which is row oil mixed with very aggressive salty underground water. But the above-mentioned sand or clay in the underground water very slowly destroys the gasket; so this salty and aggressive water may gradually enter into the motor cooling oil and slowly spoil its electrical insulation (resistance) from normal at the beginning - more than 500 kOhm, - to a risky one of less than 30 kOhm (and in a bad case - exponentially throughout one-to-three or more months).

ESP electrical motors are the most expensive parts of any ESP. To run the ESP under risky conditions is dangerous. Therefore, down-hole submersible telemetry units (system) must control its electrical resistance in real-time very cautiously. It should be mentioned here, that usually telemetry input/output signals of down-hole units are around standards of 0 – 10 V or 4 – 20 mA.

Telemetry signals are sent up from the down-hole unit to the display/control box by means of the same powerful motor supply cable between the ESP motor and the variable speed driver PWM convertor with a built-in electronic active filter (as shown in Fig. 2). The PWM convertor waveform distortion factors of output voltage and currents correspondingly

are: without filter - $K_{UOutput}$=43.73%, $K_{IOutput}$=6.75%, and after the filter is applied - $K_{UOutput}$=2.55%, $K_{IOutput}$=0.58% [3].

Theoretically, harmonics may affect such equipment in several ways but mainly:

• Notches in the sinusoidal voltage can cause malfunctions due to a misfiring thyristor or IGBT;

• Voltage harmonics may cause ignition beyond the required time;

• A resonance in the presence of different types of equipment can lead to overstrain and machines shacking.

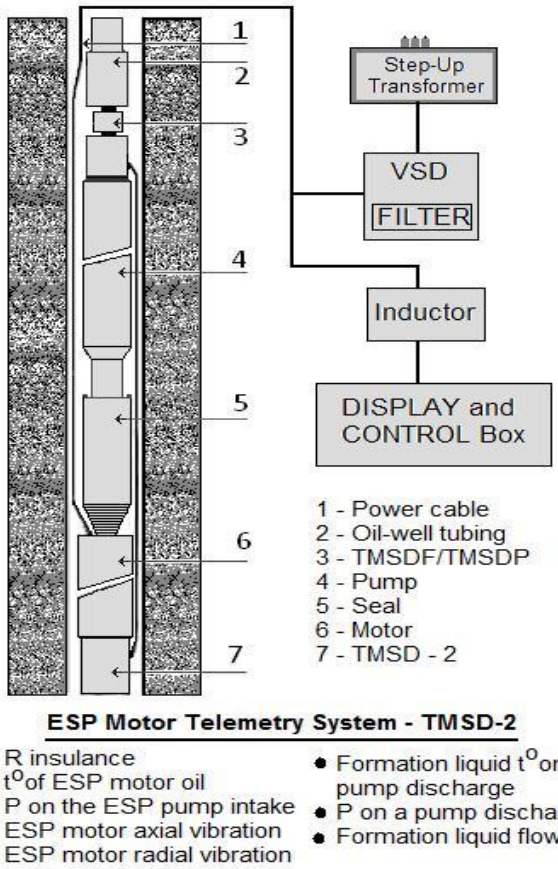


Fig. 2. The structural cart of the ESP unit equipment complex consists of a step-up transformer, display/control and inductor boxes, VSD convertor with built-in FSA and down-hole unit submersible telemetry system (TMSD Flow/TMSD Pressure or TMSD-2).

For ESP motors, these waveform distortion factors of the output voltage and current are not affecting the ESP motor-pump performance activity. But voltage and current residual

harmonics may be higher than some of the telemetry signals during some regimes and may confuse the display/control system and cause malfunctions in the equipment of the ESP (such as: data errors, failures, even short circuits in the motor, etc.). Interference induction and stray pick-up may be found comparable or even higher than the useful telemetry system signals at the moment. The harmonics in power circuits make noises in the chain of the telemetry and control lines. This small noise leads to a certain discomfort, but as it increases, the transmitted information misinterpretation may appear; or it will decrease and in limited cases, the telemetry becomes impossible in totality. In the case of any technological changes in the electrical and telemetry, there should be considered the impact on telemetry lines, and importantly, on the controlled equipment.

All of these regime parameters controlled by ESP telemetry (R–resistance, t_o – temperature of the motor cooling oil, motor vibrations and pump technological parameters) are very important for the ESP normal exploitation. But from the very sensitive ESP electrical motor viewpoint, the most important parameter of the ESP motor is the resistance R of its cooling oil, which depending on the sand factor, the gasket may wear out and provoke the leaking of salty underground liquid through the gasket and into the motor. So, the resistance R of the cooling oil starts to slowly decrease exponentially throughout three-to-five months, or even more, from normal - more than 500kOhm, - to risky – fewer than 30kOhm,- and its telemetry voltage signal proportionally and slowly decreases from 10V to 0.6V. This signal too, as all ESP telemetry signals, is transmitted from the down-hole unit upward to the display/control box of the ESP using also the same powerful motor feeding cable. This cable fills up by interference induction from currents, internal and external networking communication stray pick-ups, VSD convertor powerful PWM pulse interference and residual after-filter harmonics such as: random, casual or incidental noise [10, 11].

Thus, this paper deals with an attempt to elevate the accuracy in the continuous interpreting of the R-signal from the above-mentioned corruptions by jams of interferences, harmonics and noises through the help of well-known modern adaptive methods. Any error in the interpretations of the R-signal may bring the wrongful early prediction of the critically low R-signal and the unreasonable expensive lift of the ESP for the motor cooling oil removal and its renewal service. Or in the worst case of late prediction, may cause a short circuit inside the motor leading to an emergency lift of the ESP for service and a more expensive restoration of the motor.

2.2. Interferences and Noises in ESP Telemetry

Thus, the jam of interferences and noises in all ESP telemetry, which accompany the useful telemetry signals, consist of: electromagnetic interference induction from cable currents, internal and external network communications stray pick-ups, VSD converter powerful pulse intervention and residual after filter harmonics: random, casual or incidental noise, etc. Analyzed were several typical for VSD with PWM converters harmonics spectra for the steady state regime of ESP and the typical share and fraction of harmonics values after FSA filters are presented for this case [3] as noted in the below figures:

Usually used to assess the impact of the various harmonics, the coefficient representing harmonics taken with certain weights will be used here. The most common are two factors: photometric weighting and C-transfer [12, 13, 14].

3. AN OBJECTIVE AND COMPARATIVE EVALUATION OF MATLAB NOISE-CANCELLING ADAPTIVE METHODS

For the comparative evaluation of MATLAB noise-cancelling (filtering) adaptive methods, here applies a twice heavier case for the study – voltages of the jam of accompanying harmonics V_H =10V each (not 3 – 5V) and random noise V_{RN} =1V (not 0.1 – 0.5V). As mentioned above, the controlled and very useful exponential variable – the parameter of the ESP motor cooling oil R–resistance signal can decay from 10V to 0.6V. All of the MATLAB Adaptive Filtering Methods presented below were one-by-one tested under the above-mentioned conditions for R–resistance decreasing exponentially the signal, corrupted by the jam of the accompanying useful signal harmonics and noises. Afterwards, the R–resistance signal is recognized and analyzed [15 - 18].

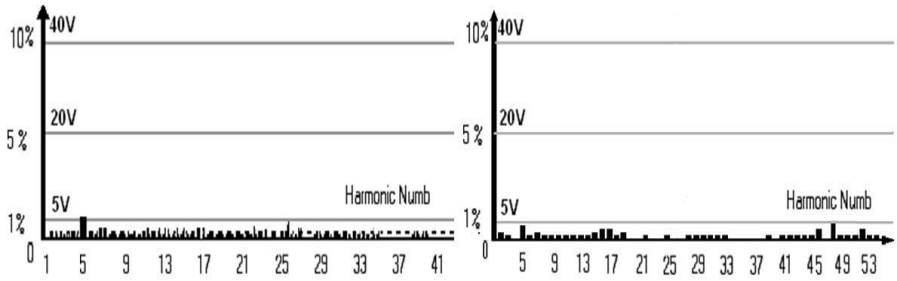


Fig. 3. The typical share of current (a) and voltage (b) harmonics values of the VSD PWM converter after build-in FSA.

a. 10-20th and 40-50th current harmonics of ~0.5-1.0% (~2-4A) of I_{Peak} value (374A) of the motor current,
b. 10-20th and 40-50th voltage harmonics of ~0.5-1.0% (~3-5V) of the motor voltage V_{Peak} value (496V).

Analysis of Results

Thus, the signal types are: exponential, sinusoidal and random. The adaptive filter algorithms were applied to observe the value of the exponential variable parameter of the ESP motor cooling oil R(t)– resistance signal - corrupted by the jam of accompanying interferences and noises. All methods were tested under a mixture of random noises and dominated harmonics for *f=1; 2.5; 5 kHz.*

There are two main zones in the below result curves:

1.The first zone - filter output signal at the beginning of the filter adaptation time T_{AD} = 3 – 30 days (not very important) which converge towards the desired exponential R(t)-signal and then continuously controls it until the critical point.

2.The second zone - filter output signal at the end of the observing time – three and more months (up to a year), - till the very important critical point, when signal R(t) ≤ 0.6V,

which means that the ESP motor cooling oil resistance less than 30 kOhм – it is an extremely risky moment and the ESP must be switched off (from the maintenance instruction). It should be reminded here once again that any error in the interpretation of the R(t)-signal critical point may bring a wrong and unreasonably early expensive lift of the ESP for the motor cooling oil removal and renewal service, or in the worst case of the R(t)-signal critical point's late prediction - to short a circuit inside the motor, causing an emergency lift of the ESP for service and a very expensive restoration of the motor.

Unfortunately the characteristics of some of the adaptive filter algorithms have shown very low frequency ripple oscillation (like *Sign-error, Sign-data and Sign-sign filters*), which may bring additional errors in the late or early interpretations of the critical point issue. The less amplitude of this oscillation, the better the adaptive filter algorithm (method). Recognized as the best was - *The Normalized LMS FIR filter algorithm* adaptfilt.nlms. One of the best approaches to avoid this very low frequency ripple oscillation is to add a block of Wavelet filter algorithm into the just mentioned best adaptive filters algorithms too. This was done at the end of study and improved results as well [19, 20 and 21].

Matlab Adaptive Filtering Methods And Their Results' Analyses

Table 1. Matlab Adaptive Filtering Methods and their results' analyses

Type of Adaptive Filter Methods	**Harmonics Frequency (kHz)**	**Remarks**
1. The Normalized LMS FIR filter algorithm (adaptfilt.nlms)	1; 2.5; 5	The Best
2. The Sign-data LMS FIR filter algorithm (adaptfilt.sd)	2.5	Good
3. The Sign-error LMS FIR filter algorithm (adaptfilt.se)	2.5	Good
4. The Sign-sign LMS FIR filter algorithm (adaptfilt.ss)	2.5	Good
5. The Traditional LMS FIR filter algorithm (adaptfilt.lms)	2.5	Fair
6. The Delayed LMS FIR filter algorithm (adaptfilt.dlms)	2.5	Fair
7. The Adjoint LMS FIR filter algorithm (adaptfilt.adjlms)	2.5	Very Bad
8. The FFT-based Block LMS FIR filter	2.5	Very Bad

algorithm (adaptfilt.blmsfft)		
9. The Filtered-x LMS FIR filter algorithm (adaptfilt.filtxlms)	2.5	Very Bad
10.The Block LMS FIR adaptive filter algorithm (adaptfilt.blms)	2.5	Very Bad

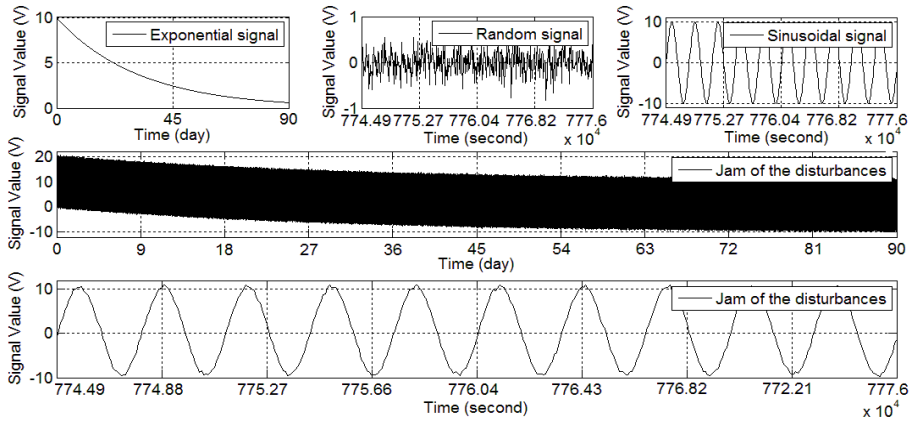


Fig. 4. Signals: useful exponential, which is corrupted by a jam of sinusoidal harmonics and random noises.

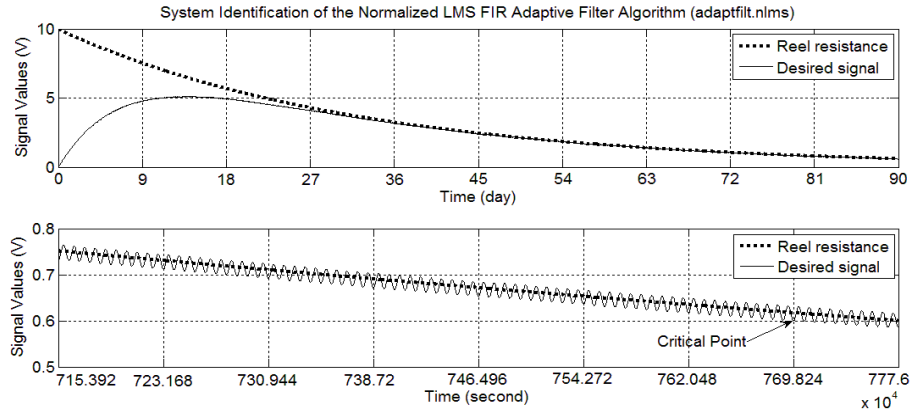


Fig. 5. The Normalized LMS FIR adaptive filter algorithm adaptfilt.nlms (1kHz)

(The Best)

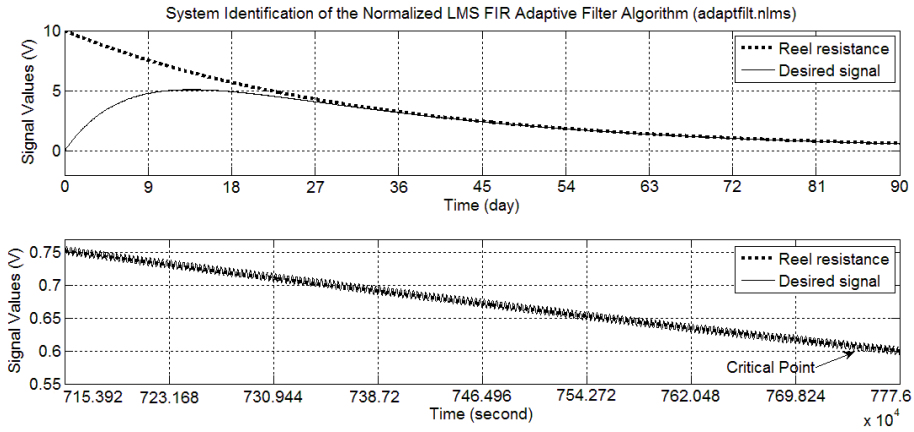


Fig. 6. The Normalized LMS FIR adaptive filter algorithm adaptfilt.nlms (2.5kHz) (The Best)

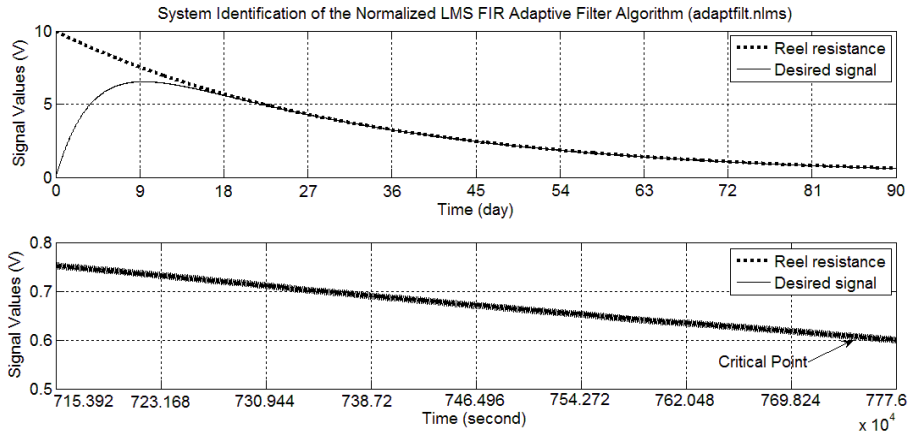


Fig. 7. The Normalized LMS FIR adaptive filter algorithm adaptfilt.nlms (5kHz)

(The Best)

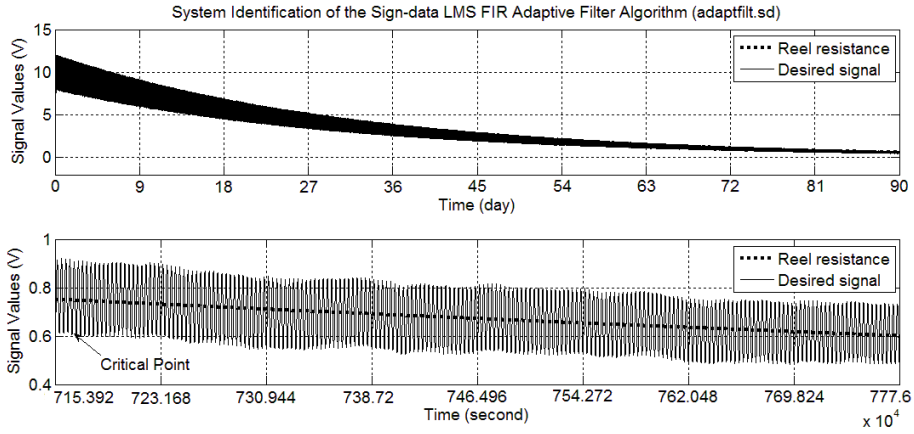


Fig. 8. The Sign-data LMS FIR adaptive filter algorithm adaptfilt.sd (2.5kHz)

(Very good)

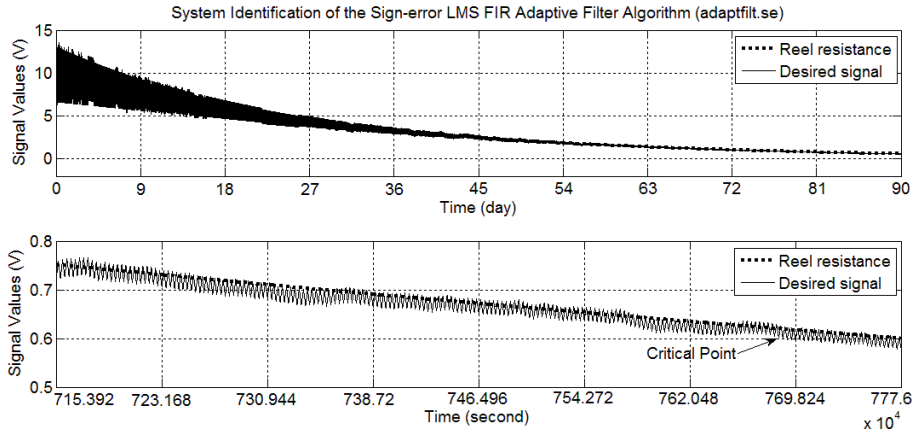


Fig. 9. The Sign-error LMS FIR adaptive filter algorithm adaptfilt.se (2.5kHz)

(Very good)

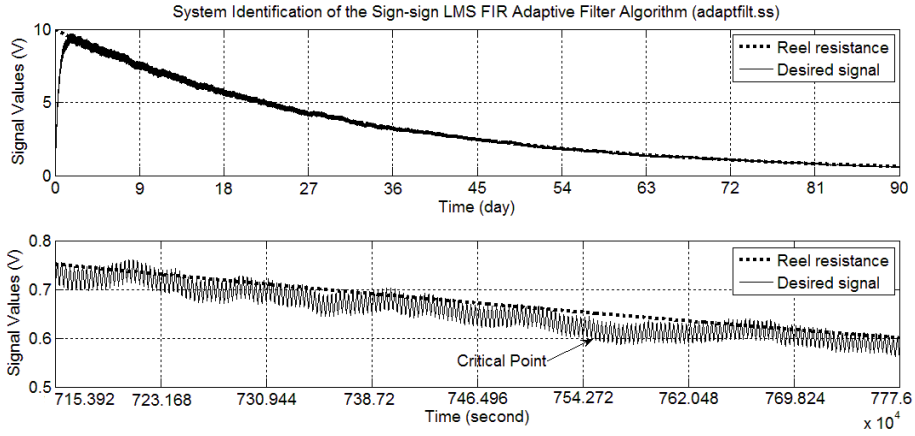


Fig. 10. The Sign-sign LMS FIR adaptive filter algorithm adaptfilt.ss (2.5kHz)

(Very good)

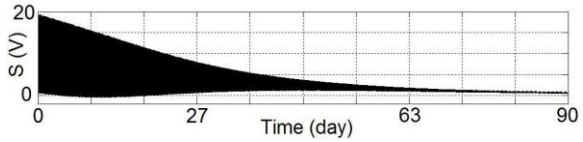


Fig. 11. .The Traditional LMS FIR adaptive filter algorithm adaptfilt.lms

(2.5kHz) (Fair)

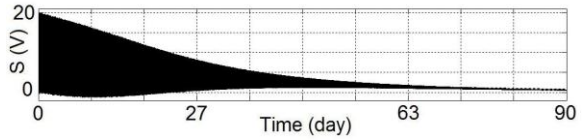


Fig. 12. The Delayed LMS FIR adaptive filter algorithm adaptfilt.dlms (2.5kHz) (Fair)

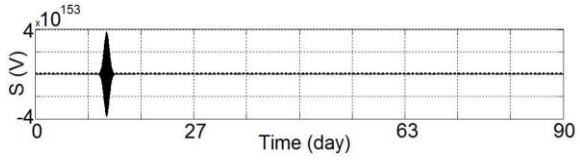


Fig. 13. The Adjoint LMS FIR adaptive filter algorithm adaptfilt.adjlms (2.5kHz) (Very Bad)

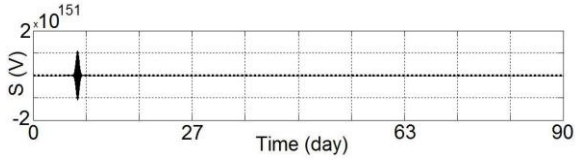


Fig. 14. The Block LMS FIR adaptive filter algorithm adaptfilt.blms (2.5kHz) (Very Bad)

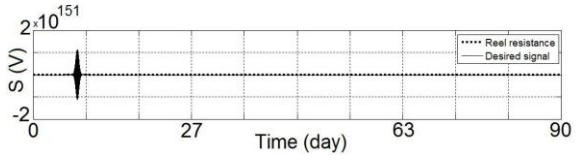


Fig. 15. The FFT-based Block LMS FIR adaptive filter algorithm adaptfilt.blmsfft (2.5kHz) (Very Bad)

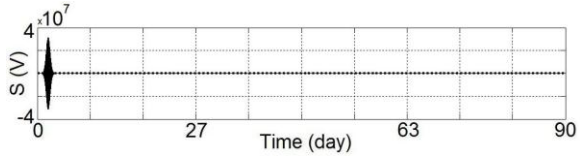


Fig. 16. The Filtered-x LMS FIR adaptive filter algorithm adaptfilt.filtxlms (2.5kHz) (Very Bad)

4.CONCLUSION

In this paper, several Adaptive Filter algorithms from MATLAB have been applied to solve in real time the problem of early prediction of disruptions in the oil industry Electro Submersible Pump (ESP) motor. From the analysis of the results, it is possible to claim that the start of trouble is predictable within a very long time interval of practical interest. Unfortunately, some of the adaptive filter algorithms have shown bad (4) and fair (2) results, which may bring additional errors in the late or false early interpretation of the critical point of the ESP motor insulation disruption issue. Some of the adaptive filter algorithms have

shown successful and very good results of the early prediction of the ESP motor real insulation disruption (like Sign-error, Sign-data and Sign-sign filters). The best among the ten analyzed adaptive filter algorithms (methods) for application in ESP telemetry was recognized as - The Normalized LMS FIR filter algorithm - adaptfilt.nlms.

6. REFERENCES

1. F. Mohn, Electric Submersible Pumps. Bergen, Norway: AS Oil & Gas, Division, *Frank Mohn AS*, 1994.
2. William C. Lyons, Gary J. Plisga, Standard handbook of petroleum and natural gas engineering, *Gulf Professional Publishing*, 2005.
3. Smart tools for oil industry, The leading manufacturer of the oil surface equipment industry in Russia, *www.electon.ru*.
4. Shynk J.J., Frequency-Domain and Multirate Adaptive Filtering, *IEEE Signal Processing Magazine*, Jan. 1992, vol. 9, no. 1, pp. 14-37.
5. Hayes, M., Statistical Digital Signal Processing and Modeling, *New York, Wiley*, 1996.
6. Alexander D. Poularikas and Zayed M. Ramadan, I. K., Adaptive Filtering Primer with Matlab, *International Pvt. Ltd*, January 2006.
7. P. G. Ali-Zade, P. K. Kulizade, Method of electro submersible pump motor insulation checking, "*Patent USSR, AC SU 271644*", 1968/1970.
8. IEEE Recommended Practice for Specifying Electric Submersible Pump Cable-Ethylene-Propylene Rubber Insulation, *IEEE Std 1018™, IEEE Industry Applications Society, New York*, NY 10016-5997, USA, 13 April 2005.
9. Dillard, Steve M., Greiner, Thomas D., Transient Voltage Protection for Induction Motors Including Electrical Submersible Pumps, *IEEE Transactions on Industry Applications*, March 1987, vol. IA-23 Issue:2, pp. 365 – 370.
10. Boldea, Ion, and S.A. Nasar. Electric Drives. Boca Raton, *Florida: CRC Press*, 1999.
11. Bose, B. K. Modern Power Electronics and AC Drives. *Upper Saddle River, NJ: Prentice-Hall*, 2002.
12. International Telegraph and Telephone Consultative Committee (CCITT), *www.itu.int/net/about/history.aspx.*
13. Bell Telephone Company and Edison Electrical Institute, *www.edisonfoundation.net/iee/reports/index.htm.*
14. Inter branch rules on a labor safety (rules) at operations of electro-installations. ПОТ РМ-016-2001, PEEP and PUE (in Russian), *http://www.enerkomp.ru/documents/need.html.*
15. Moschner, J.L., Adaptive Filter with Clipped Input Data, *Ph.D. thesis, Stanford Univ.*, Stanford, CA, June 1970.
16. Gersho, A, Adaptive Filtering With Binary Reinforcement, *IEEE Trans. Information Theory*, March 1984, vol. IT-30, pp. 191-199.
17. C. F. N. Cowan and P. F. Adams and Peter M. Grant, Adaptive Filters, *Prentice Hall Signal Processing Series*, Englewood Cliffs, 1985.
18. Wan, Eric., Adjoint LMS: An Alternative to Filtered-X LMS and Multiple Error LMS," *Proceedings of the International Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 1997, pp. 1841-1845.
19. S. Mallat, A Wavelet Tour of Signal Processing. *Cambridge: Academic Press*, 1999.
20. M. Yilmaz, R. N. Tuncay, and O. Ustun, A wavelet study of sensorless control of brushless DC motor through rapid prototyping approach, *IEEE International Conference on Mechatronics*, ICM 2004.

21. M. Yilmaz, R. N. Tuncay, O. Ustun, T. P. Krein, Sensorless Control of Brushless DC Motor Based on Wavelet Theory, *Electric Power Components & Systems, Taylor & Francis*, October 2009, volume 37, number 10.
22. Consider the silent treatment, Noise control in Alberta, Pipeline & Gas Journal, April 2012, Vol. 239 No. 4

7. APPENDIXES

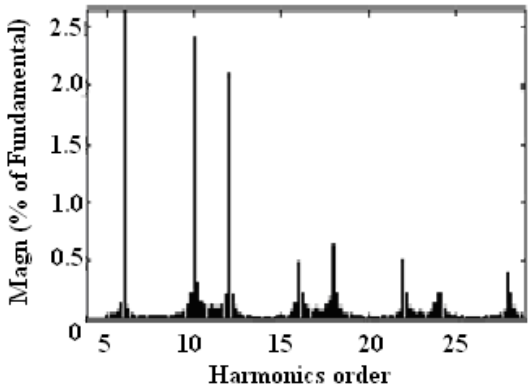


Fig. 17. German-Galkin C. G., Spectral analyses of processes in power semiconductor converters in MATLAB package, in Transactions, "Corona Print", 2001, 315 pp.

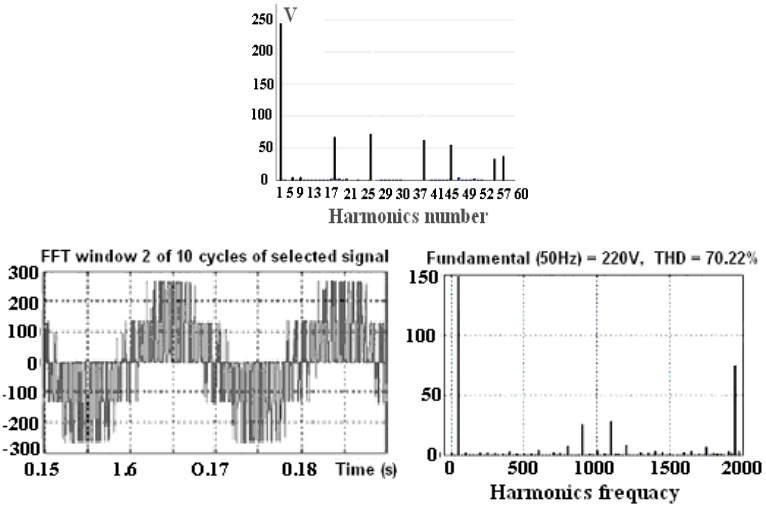


Fig. 18. www.mathworks.com, www.matlab.ru, www.exponenta.ru

Technological Prototype of The National Wetlands Inventory - A Multidisciplinary Development

Lucia Patricia Carrillo Velázquez

Universidad Nacional Autónoma de México
(National Autonomous University of Mexico)
CEIICH, PAPIME 304909

carrillopatricia4@gmail.com

Abstract: In Mexico, many wetlands face serious deterioration and over-exploitation due to mining and pollution of their waters, threatening their natural functions as potential water reserves for different human uses. That is why this deterioration is considered a national priority that requires identification and characterization of Mexican wetlands to create strategies and policies for their preservation, protection, restoration and rational use.

The multifactorial and national character of this phenomenon requires multidisciplinary analysis and the use of technological resources to process information on the organization, operation, time and space, a mong others, under special conditions.

This paper emphasize the technological integration into the multidisciplinary group and describes a proposal to develop a technological platform through a computer-based prototype, based on a methodological model of software engineering that makes explicit the conceptual design vis-à-vis the difficulty of demonstrating its usefulness and functional structure in a way that can be understood by all disciplines involved through a software prototype.

Keywords: Telematics, multidisciplinary approach, wetlands, GIS.

Wetlands in Mexico: A Priority

The recognition of wetlands in Mexico is a national priority since:

> "In Mexico, many wetlands face serious deterioration and over-exploitation due to mining and water pollution .... With the increasing demand for water for different human uses, the wetlands, as existing and potential reserves of drinking water, will increasingly be reduced and dried, unless strategies and policies to identify, define, preserve, protect, restore and use their numerous biotic resources rationally are implemented. [That is why] it is a priority to implement policies and programs, under the provisions of the National Water Law that requires an information platform on the country's wetlands, which particularly include their location and ecological aspects.

These studies will result in a better understanding of their structure and functions as a component of the hydrological cycle, and in the context of a strategic vision, will make it possible to create proposals and set priorities that support the comprehensive management of national waters." [6]

Fundamental Requirements to Develop a National Wetland Inventory

The multifactorial and national character of this phenomenon requires multidisciplinary analysis and the use of technological resources to process information on the organization, operation, time and space, among others, under special conditions. Hence in developing a National Wetlands Inventory (NWI), three basic requirements need to be considered:

1. **The methodological process:** In order to carry out the required identification, characterization, and definition it is fundamental and essential to have an analysis criteria system that guides and facilitates the recognition of patterns, variables and decision-making criteria, among others, so as to create software information models. Especially to identify, characterize, monitor and evaluate the Mexican wetlands with the purpose of generating knowledge on water that explains the problems and facilitates the creation of strategies to attend to the needs of the sector by identifying and taking advantage of such opportunities. [6] There are different classification models, notably that of Ramsar, [1] a classification system proposed by the international community during the Ramsar Convention [9] with the purpose of identifying wetlands of international importance.

   However, those models, and especially Ramsar's limit the recognition of specific variables and patterns unique to Mexico, due to such models' intentions of standardization. It is thus necessary to integrate an analysis system based on the needs and specificities of our country as the foundation for the construction of explicit knowledge of Mexican wetlands and their representation in a national inventory.

2. **Collaborative and multidisciplinary work:** the detailed and specific characteristics required by the integration of a wetlands inventory in our country implies the convergence of a diversity of fields of knowledge related to the characterization of water bodies. It therefore requires the integration of a multidisciplinary instrument to gather information, initially "managed" by a technological systematization process, which will later be interpreted. This integration reveals the need to address the actions of identification and definition of the boundaries of the wetlands via a collaborative multidisciplinary strategy to ascertain the level of deterioration and overexploitation due to mining and contamination with the goal of making explicit the knowledge that is needed to preserve, protect, restore and use wetlands rationally.

3. **The technological factor:** at first sight, the usefulness of a system based on Information and Communications Technology (ICT) to define and geographically locate the wetlands is evident. This responds to the need to recognize and locate wetlands.

   In Mexico, we can observe the prevalence of these types of systems in earlier experiences discussed in congresses. Such systems can be characterized as Geographic

Information Systems (GIS), since they systematize and process data whose geographic reference is shown in digital maps.

That is why we also consider, for the proposed design, the perspective *for Technological use* if *commercial or open-source software* is used to implement a technological platform for the NWI and having to "adjust" in such case our needs to the usability limitations according to the terms of operation of such *software,* which without further analysis suggests that attention to the proposed needs would limit the *software*'s capabilities. Thus, we propose the option of implementing knowledge creation and transformation processes that translate into technological development or appropriation capabilities aimed at supporting the dynamic processes of a function structure *ad hoc* to the integration of an NWI in Mexico.

The analytical integration of these factors led us to these basic questions: How to articulate the knowledge-building process in order to make the confluence of different disciplinary perspectives regarding the explicit knowledge on wetlands in an NWI operational? And how to articulate the process required for building knowledge between the multidisciplinary group and the technology experts group which is necessary for the optimal and adequate incorporation of technological resources?

Technological Integration into the Multidisciplinary Group

Understanding that multidisciplinary integration results from a non-linear process of knowledge creation, transformation, explicit representation and communication, which takes place in a cognitive dimension involving different actors, it is considered essential to establish the relevance of technology in selecting the instrumental resources that "manage" this process and therefore determine its operational quality. We believe that implementing a GIS with the specific purpose of defining and mapping wetland location dismantles the synchronicity of the multidisciplinary knowledge-building process that can result in an instrument that separates the explicit representation of knowledge from the essential communication process required for the interaction among different experts, especially those from the technology field.

In particular, we are referring to the conception of technology that guides structure instrumentation and comprehensive operation of a multidisciplinary research team whose activities are directed at a common object of knowledge. In the case at hand, the goal is the integration of an inventory to identify, define, characterize, evaluate and monitor the country's wetlands.

In earlier research conducted with this perspective, we have had successful experiences when aspiring to the level of technological appropriation and development that put into practice processes of technological management of multidisciplinary knowledge with the engagement of experts in technology in the research team. Ongoing and repetitive academic dialoguing encourages interaction among disciplines [3] that results in the reformulation of concepts, methodologies and dimensions of analysis in which the technological perspective becomes tacit and fundamental, thereby developing the capacity to identify and specify technological requirements. These actions are essential in software's fundamental phase of development, or lifecycle in order to ensure that it complies with the requirements of required application and validates its development procedures, ensuring that

the software methods and techniques that are to be used are appropriate. All this is set forth from the *engineering* perspective, which studies and proposes *software* development and maintenance methods and techniques.[7]

The capacity to design and innovate technological developments in research groups, especially in development of application software or of models for technology operation or processing are thus created.

Considering the above as a research assumption, the participation of a multidisciplinary team in technological design is an indicator of appropriation that will be verified by the same team according to how useful technological output results and to the extent to which it meets their needs, in this case a software system as an explicit representation of scientific knowledge with the double function of research resource and multidisciplinary knowledge "manager." In the particular case of this empirical experience, it implies a technological platform for software processing of multidisciplinary knowledge of the Mexican wetlands and a technological resource that makes the interaction between experts in building multidisciplinary analysis criteria to identify, characterize, monitor and evaluate these wetlands both realistic and operational. The approach underlying the experience that is documented here has been described in these terms.

A Technological Prototype for the National Wetlands Inventory

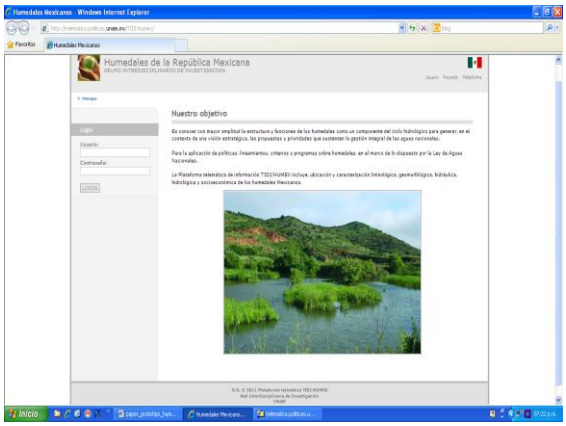
From the software engineering field, a preliminary outcome of multidisciplinary interaction is a prototype developed in congruence with the Rapid Application Development methodology, in which software development involves minimal planning, but has the advantage of immediate operation through prototypes. [8]. The use of this methodology was useful for the dynamic nature of the design process conducted by the team during the construction of a logic of multidisciplinary significances and the production and simultaneous transformation of the technological system's components required for the explicit representation for this constructive process.

For their part, the integration of websites using a Graphic User Interface (GUI) to link the component of the System operating the Database Management System (DBMS) and the processing of analysis models is essential since it allows for both the creation of a context of virtual interaction of the reality of wetlands in Mexico that is required to provide significance to the representations of scientific knowledge. (see Figures 1 and 2).

Figure 1

TIDI-HUMEX Platform

National Wetlands Index



Source: Carrillo Velázquez Lucia Patricia

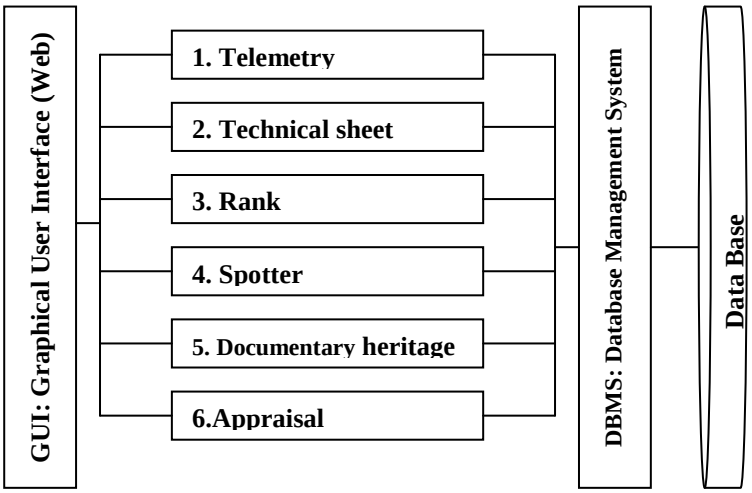
http://telematica.politicas.unam.mx/TIDI-humex/
[Online, consulted in Jan. 2011]

Figure 2

TIDI-HUMEX Platform

National Wetlands Index

Fundamental Processes for Multidisciplinary Research



Source: Carrillo Velázquez Lucia Patricia

Proyect: 2009. National Wetlands Index

Final considerations

The prototype resulting from a technological platform currently in process for the creation of a NWI in Mexico prompts us to consider the basic and traditional criteria of a system's value and of its software engineering, namely quality vs. low cost, reduction of the risk of needs not being met –including commercial needs– and delivery deadlines. In this project, we assess the quality of a technological system and its software engineering with the Joint Application Development methodology [7] at a public university in Mexico. In its incidence level, the project manifests as a resource to manage scientific knowledge, specifically in its technological appropriation and innovation capacity developed by researchers from different disciplines as an explicit representation resource that functions both as a research tool and as a resource that enhances the intellectual capital of an academic institution.

References

1. Berlanga-Robles, César Alejandro; Ruiz-Luna, Arturo; Lanza Espino, Guadalupe. *Esquema de clasificación de los humedales de México*. Investigaciones Geográficas, Num. 66, August, 2008, pp. 25-46. Universidad Nacional Autónoma de México. Mexico City. [Accessed in January 2010]. http://redalyc.uaemex.mx/src/inicio/ArtPdfRed.jsp?iCve=56911123003
2. Botkin, John C. *Customer Involved Participation as Part of the Application Development Process.* n.d. [Accessed in February 2011].http://wwwsgi.ursus.maine.edu/gisweb/spatdb/amfm/am94001.html.
3. Carrillo V., L. Patricia. Semantic *Web and Logic of Significances in the Knowledge Management of a Social Science Research.* V Conference Athenaeum 2011.Thricky, India.
4. *Proyecto PAPIME PE304909.* CEIICH-DGAPA, UNAM. 2009.
5. *Sociedad del conocimiento. Academia, complejidad, administración y tecnología*. SITESA-UNAM. D.F. México. 2008.
6. Consejo Nacional del Agua (CONAGUA) and Consejo Nacional de Ciencia y Tecnología (CONACYT). 2008. Convocatoria 2008/1. *Fondo Sectorial de Investigación y Desarrollo sobre el agua.* CONAGUA-CONACYT. Mexico City.
7. Dennis, Alan R.; Glenda S. Hayes, and Robert M. Daniels, Jr. Business Process Modeling with Group Support Systems. *Journal of Management Information Systems* 15 (4): 115–142. 1999. [Accessed in February, 2011].http://www.jmis-web.org/articles/v15_n4_p115/index.html
8. Gottesdiener, Ellen. RAD Realities: Beyond the Hype to How RAD Really Works. Application Development Trends. 1995.
9. Ramsar, Convención de Ramsar**,** Secretaría de la Convención de Ramsar. *Manual de la Convención de Ramsar. Guía a la Convención sobre los Humedales.* (Ramsar, Iran, 1971)**.** 4a. edición. Secretaría de la Convención de Ramsar. Gland, Switzerland. 2006.

Agents Knowledge Models for Situation Management Systems

Oleksiy E. Kovalenko

Institute of Mathematical Machines and Systems Problems of NAS of Ukraine, Visual Information Processing Systems Theory and Practice Department, Glushkov Av. 42,
03680 Kyiv, Ukraine
koval@immsp.kiev.ua

Abstract. Situation management systems are important components of up-to-date information technologies for business and public management. Recent trends in complex information systems creation are widely make use of agent-oriented programming, knowledge bases, cloud computing and so on. Issues of knowledge models designing for program agents are discussed. The specifics of situation management systems are described. The aspects of general and special knowledge of agents are outlined. Knowledge models for program agents of different tiers of situation management systems are proposed.

Keywords: agent-oriented system, knowledge model, ontology, situation management system.

1 Introduction

Situation management systems based on information technologies are used in various branches of human activity [1, 2]. Situation management (SM) is considered "as a framework of concepts, models and enabling technologies for recognizing, reasoning about, affecting on, and predicting situations that are happening or might happen in dynamic systems during pre-defined operational time" [2]. The Multi-Agent System (MAS) [3] due to its characteristics directly fit for the solving SM problems [4]. MAS can be used as supporting tools for organizing and serving integrated environments of SM. Such type of SM environment we call as Agent Based Environments of Situation Management (ABESM).

SM is a complex multilayered technology with variety of the interrelated tasks on each layer. Appropriate agents on various layers perform the functions of dispatching, coordination, services seeking and providing, security etc. Agents during functioning must use the appropriate knowledge. Knowledge of an agent is a fragment of knowledge domain of SM problem. Knowledge domain of SM problem is an element of SM model.

2 Situation Management Model

SM model includes submodels for representing different aspects of SM activity realization. These submodels are Situation Management Structure, Situation Management Ontology, Situation Management Workflow. Integration of submodels provides completeness and adequateness of SM process.

2.1 Situation Management Structure

SM consists of some essential stages: situation detection, situation identification, situation control goal-setting, resource definition, constraints definition, control planning, affecting on the situation, situation state monitoring and perception, plans correction, completion, summing-up. These stages represent the structure of SM. Structure elements define the set of concrete services required for specific phase of SM process. Services are provided by suitable delivery agents. Agents use appropriate knowledge from SM knowledge domain for services identification and mutual coordination. The SM knowledge represents as ontology.

2.2 Situation Management Ontology

SM ontology represents knowledge for appropriate domain and brings together situation and management ontology. The situation ontology depicted on Fig.1 [5].

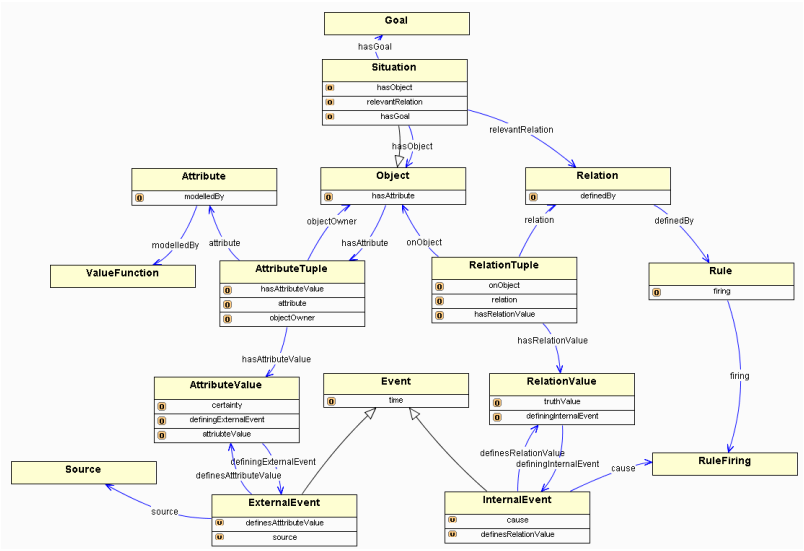


Fig. 1. Situation Core Ontology.[5]

SM in general is the business process of project management dealing with planning, organizing, securing, leading, and controlling resources to achieve specific goals. The view of business process ontology presented on fig. 2 [6]. The meta-model of the business process ontology can be defined in a way that suits the domain expert, the requirements engineer and the business process modeller, who may not be an IT expert. A business process ontology would describe all concepts related with a business process. In particular, it would define entity types such as business activity, business document, business object, business event,

business rule, role, resource and control flow. In addition, relationships among entities are defined.

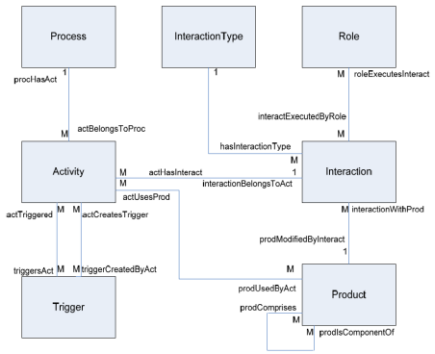


Fig. 2. Business Process High-level Ontology.[6]

The real value of a business process ontology lies in the ability to extend the software development value chain. The combination of business process management and semantic technologies leads to Semantic Business Process Management (SBPM). SBPM based on top level ontologies: event ontology, situation ontology, process ontology, temporal ontology, spatio ontology, specific domain ontologies, task ontologies, and application ontologies [7]. SM business process realized thru its workflow.

2.3 Situation Management Workflow

SM workflow would be based on standard model. Such model is Workflow Reference Model proposed by The Workflow Management Coalition Specification [8].

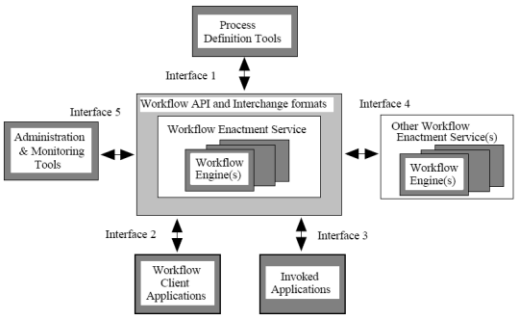


Fig. 3. Workflow Reference Model. [8]

All workflow systems contain a number of generic components which interact in a defined set of ways; different products will typically exhibit different levels of capability within each of these generic components. To achieve interoperability between workflow products a standardized set of interfaces and data interchange formats between such

components is necessary. Standardized interfaces and data interchange formats provides by workflow services. Discovery and delivery of services are provided by agents with knowledge about SM. Example of the workflow ontology can be downloaded from [9].

3 Agents Knowledge Models

Agent's behavior is determined by its internal knowledge and behavior model. The most common model of agent behavior is the BDI model supported by the FIPA specification. Knowledge of the agent to determine its behavior include knowledge about the environment, resources (services), the mechanisms of communication and coordination.

3.1 Agents-Based Hierarchical Environment Model

According to SM model the agents distributed over the levels of hierarchy to support the functions of its level. Agents are software entities that implement a model of behavior based on internal knowledge representation in a formalized manner. The interaction of agents of the same level based on the principles of choreography, and the agents of different levels - on the principles of orchestration.

The level of the agent hierarchy determines its knowledge, functionality, language tools and other facilities. Whereas the agents in agent-based SM are used for services discovery and delivery so theirs languages hierarchy must meet the hierarchy of web services stack languages. The web-services stack languages proposed in [10] depicted on fig.3.

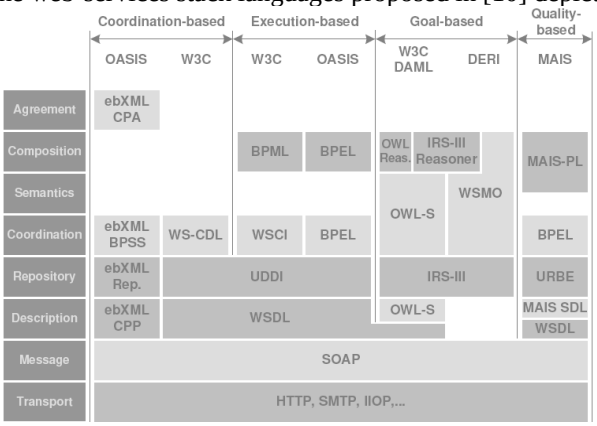


Fig. 4. Web Service composition-oriented protocol stack of vendor-specific and standardized protocols and languages. [10]

In recent version the web services protocol stack was expanded by xPDL 2.2 specification [11], and BPML was replaced by BPMN (Business Process Model and Notation) [12] and WS-BPEL (Web Services Business Process Execution Language). The BPMN specification includes an informal and partial mapping from BPMN to BPEL [13].

3.2 Agent Communication Model

To perform its functions agents have to communicate with each other and the services. A generic communication stack (CS) for agent communications is shown in Figure 1 [14] and logically corresponds to the abstract communication architecture standardized by FIPA in [15].

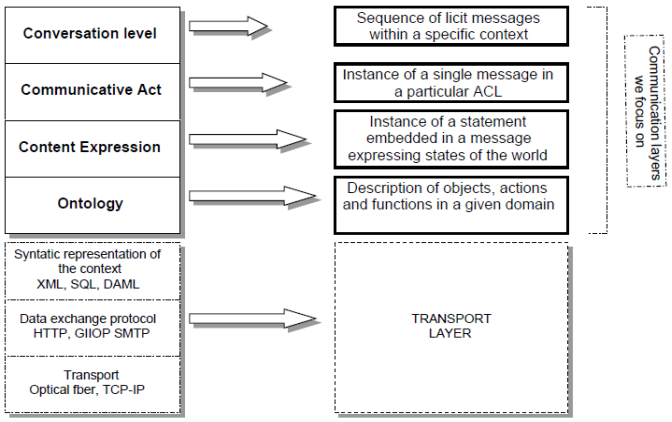


Fig. 5. A generic communication stack for agent communications.[14]

So for communication with services agents contain knowledge about services interfaces in format meeting the requirements of WSDL [16]. For Java agents specification is Java APIs for WSDL (JWSDL) [17].

3.3 Agent Ontology

Agent ontology is the agent of knowledge and provides adequate model behavior in SM runtime environment. To ensure consistency and completeness the agents knowledge is a relevant piece of knowledge domain knowledge SM. So the agent ontology consists of appropriate ontology fragments from SM business process ontology, SM workflow ontology, specific ontology of situation domain, ontology of environment, descriptive ontology of services and agents, communicative languages and protocols, etc.

Specific ontology of agent allocated from the generic ontologies by generating and processing the request, for example with use of language OWL2 DL. [18].

3.4 Agents Orchestration

An orchestration defines the sequence and conditions in which one program agent or service invokes other program agents or services in order to realize some useful function. Thus, an orchestration is the pattern of interactions that a program agent or service must

follow in order to achieve its goal. Orchestration is used for coordination of agents from different layers of SM business process. So, the appropriate method of orchestration can be chosen from one of orchestration patterns [19] using the query to ontology. Formal representing of the orchestration may use the orchestration language [20].

3.4 Agents Choreography

Choreography is the definition of the sequences and conditions under which multiple cooperating independent agents exchange messages in order to perform a task to achieve a goal state. Choreography is used for agent coordination of the same layers of SM. Choreography, in general, is unsynchronized process. So, synchronization in agents choreography needs using of formal methods, for example choreography language [20].

Thus, the orchestration and choreography languages are logic models and may be translated into OWL DL (OWL2 DL).

4 Conclusions

The business process model and workflow of agent-based SM need integration of various methods, technologies and frameworks for adequate representing of situation and management context in supporting information systems.

These models are SM model, SM workflow model, SM ontology, hierarchical agents environment model, agent communication model, agent ontology. All ontologies integrated in generic ontology of SM.

Completeness and integrity of local agents ontologies provides by the use of queries to generic ontology of SM.

Workflow of SM realized thru orchestration and choreography with the use of special formal languages. These languages may be translated into OWL DL dialect.

References

1. Gabriel Jakobson, Lundy Lewis, Christopher J. Matheus, Mieczyslaw M. Kokar John Buford: Overview of situation management at SIMA 2005. In: Military Communications Conference, 2005. MILCOM 2005. IEEE, pp. 1630-1636. IEEE Press. (2005)
2. Gabriel Jakobson, John Buford, Lundy Lewis: Situation Management: Basic Concepts and Approaches. In: Popovich, Vasily V.; Schrenk, Manfred; Korolenko, Kyrill V. (eds.) Information Fusion and Geographic Information Systems 2007. LNG&C, vol. XIV, pp. 18--33. Springer, Heidelberg (2007)
3. Wooldridge, M.: An Introduction to Multi-Agent Systems. John Wiley and Sons, NY (2002)
4. John F. Buford, Gabriel Jakobson and Lundy Lewis: Multi-Agent Situation Management for Supporting Large-Scale Disaster Relief Operations. International Journal of Intelligent Control and Systems, vol.11, No. 4, 284-295. (2006)
5. Matheus, C., Kokar, M., Baclawski, K., Letkowski, J. An Application of Semantic Web Technologies to Situation Awareness. In: Y. Gil et al. (Eds.): ISWC 2005, LNCS, vol. 3729, pp. 944 – 958. Springer-Verlag, Berlin Heidelberg (2005)
6. Hughes, Robert: Project Management Process Ontologies: a Proof of Concept. In: UK Academy for Information Systems Conference Proceedings 2010. Paper 30. (2010). http://aisel.aisnet.org/ukais2010/30

7. Adrian Paschke: Rule-based Semantic BPM (2011), http://servicewave.eu/2011/files/2011/11/1_paschke.pdf
8. Reference Model - The Workflow Reference Model (WFMC-TC-1003, 19-Jan-95, 1.1), http://www.wfmc.org/reference-model.html
9. The Workflow Ontology, http://protegewiki.stanford.edu/images/5/5b/Workflow.zip
10. Florian Daniel, Barbara Pernici: Insights into Web Service Orchestration and Choreography. International Journal of E-Business Research, 2(1), 58-77 (2006)
11. Workflow Management Coalition Workflow Standard. Process Definition Interface - XML Process Definition Language. Document Number WFMC-TC-1025, Version 2.2, August 30, 2012, http://www.xpdl.org/
12. BPMN 2.0, http://www.omg.org/spec/BPMN/2.0
13. Mapping_BPMN_to_BPEL_Example, http://www.bpmn.org/Documents/Mapping_BPMN_to_BPEL_Example.pdf
14. Calisti, Monique: Abstracting communication in distributed agent-based systems. In: Proceedings of the Concrete Communication Abstractions of the Next Distributed Object Systems Workshop, ECOOP'02. (2002).
15. FIPA Abstract Architecture Specification, http://www.fipa.org/specs/fipa00001
16. Web Services Description Language (WSDL) Version 2.0 Part 1: Core Language, http://www.w3.org/TR/wsdl20
17. JSR-000110 JavaTM APIs for WSDL, http://download.oracle.com/otndocs/jcp/jwsdl-1.2-mrel2-oth-JSpec
18. OWL 2 Web Ontology Language, http://www.w3.org/TR/2009/REC-owl2-primer-20091027
19. Dragos Manolescu and Boris Lublinsky: Service Orchestration Patterns: Graduating from State of the Practice to State of the Art, http://micro-workflow.com/PDF/OrchestrationPatterns-poster.pdf
20. Nadia Busi, Roberto Gorrieri, Claudio Guidi, Roberto Lucchi, Gianluigi Zavattaro: Choreography and Orchestration: A Synergic Approach for System Design Service-Oriented Computing. In: Benatallah, Boualem; Casati, Fabio; Traverso, Paolo (eds.) Service-Oriented Computing – ICSOC 2005. LNCS, vol. 3826, pp. 228--240. Springer, Heidelberg (2005)

Integration of Information Services in Agent-Oriented Situation Management System

Oleksiy E. Kovalenko[1], Dimiter G. Velev[2]

[1] Institute of Mathematical Machines and Systems Problems of NAS of Ukraine, Visual Information Processing Systems Theory and Practice Department, Glushkov Av. 42, 03680 Kyiv, Ukraine
koval@immsp.kiev.ua

[2] University of National and World Economy, Faculty of Applied Informatics and Statistics, Dept. of Information Technologies and Communications, UNWE, 1700 Sofia, Bulgaria
dvelev@unwe.acad.bg

Abstract. One of mainlines of up-to-date information technologies is integration of information services into agile complex information systems. Those information services are mainly web-services. The agility is carried out on the base of agent-oriented approach. So, the agents perform functions of accounting and dispatching of distributed information services. Issues of integration and using information services in situation management systems are discussed. The functions of agents represented from the position of its behavior models. The means and frameworks for services integration are presented with the heeds of situation management systems specifics.

Keywords: agent-oriented system, agent behavior model, information service, situation management system.

1 Introduction

A collective decision making technology in situation centres (SC) is a multi-layered, complex, purposeful process that includes a wide class of preparatory procedures, organizational, technical nature [1]. The implementation of environments supporting such technologies can be made on the basis of information systems with the appropriate set of services provided at the request of the planner at each stage of the decision making process.

The term service refers to a set of related well-defined business functionalities that are built as software components (discrete pieces of code and/or data structures) functionalities that can be reused for different purposes, together with the policies that should control its usage. Services can be implemented on service-oriented architecture (SOA) concept and on the offspring of SOA – mashups and cloud computing.

Decision-making procedures use variety of services that support data integration, processing and representation, communication, coordination and dispatching of processes. Integration such services is a complex problem according to concrete objective.

2 Service utilization

For realization situational management in SC for implementing each process stage suitable groups services that form a hierarchy of services are used [2]:

- stage services;
- process services;
- procedure services;
- operational services;
- transactions services.

Standardization of services to be shared in the SC could be based on service-oriented architecture. Access to services can be carried out by means of enterprise systems cloud computing. The combination of Web services and agents through description needed for agent's services function in a knowledge agent. Standardization describe services carried out by means of language ebXML [3]. ISO-15000 standard describes typical components and formats of ebXML with the requirements compliance of metadata registry standard (Metadata Registry (MDR) standard) ISO/IEC 11179 [4]. Interpretation or application of the ebXML concepts is based on the recommendations ISO 15000-5 [5].

Services orchestration and choreography. Coordination of services that requested in the operation of SC is using orchestration and choreography. [6].

3 Agents Coordination

Diversity and initial uncertainty set of services needed to implement a particular model of situation management in SC, call for a wide selection functionality services, their availability, flexibility and autonomy. These requirements can be achieved based on the agent-oriented approach. Thus, situational management information system can be represented as a Multiagent System (MAS) [7].

Agents in a MAS must interact according to the decision-making procedures used in SC. Models of decision-making procedures are stored in the SC knowledge base (KB), which is replenished and refreshed during the operation process of SC with the experience of making preliminary decisions.

Thus, SC functioning has an evolutionary character. SC KB is a hierarchy of knowledge domains about various aspects of the functioning of the SC, which include both knowledge relating to specific areas in which the situational management implements and knowledges related to the functioning regulation of the SC, including knowledges about the organization of the collective decision-making process. Those knowledges are used to determine the list of required agents and their patterns of behavior and interaction in the SC operation.

In the collective decision-making algorithm implementation of each stage is supported by relevant agents groups ("colonies") in a way that the result of one colony is an input material for another activity, and eventually obtained the final result of the process of collective decision-making in a specific format. So, the decision is the result of superposition of multiple colonies functioning agents given the information field. Thus, there is a problem of formal definition of specifications of decision making on the basis of which is determined by the initial composition of the "peace agents" for a specific decision problem. Based on specifications determined by the composition of the colony, who is elected by the "population" of such interconnected agents.

Agent-manager in MAS must have a formal description of the general formulation of the problem of decision making in a query model of knowledge relevant subject area through which the selection of the set of necessary agents that implement each step in solving

specific problem. Agents operate within agent platform that provides lifecycle management agents and effective communication (messaging) between agents. Language Communication Agent (Agent Communications Language, ACL) defined a set of document FIPA, associated with ACL [8]. Actually, the ACL described in the document [9].

The mechanism that implements communication between agents called Message Transport Service (MTS) in FIPA documents. The document [10] identifies components that implement services and agents search, which include system management agents (Agent Management System, AMS) and directory services (Directory Facilitator, DF). AMS component is responsible for the management and implementation of white pages service that contain a list of agents registered on the platform. It is also responsible for maintaining the life-cycle of agents. DF component provides for agents yellow pages service. Yellow pages service provides registration agents services and search agent for its service. Platform agents can subscribe to DF-agent for information about registration required service. Set of yellow pages located on different agent platforms instances implement distributed yellow pages service. So the information base for the functioning of agent environment is supported.

When configuring the situational management environment the stages implementation agents realize the formation of agents set for the implementation process, the processes agents - the formation of a set of procedures agents, etc. to determine the required set of transactions for each decision-making operations.

Thus at each level is formed only information base about agents without their activation and binding in the interaction. Thus Agent Information environment formed. Further, according to the situational management technology it is necessary to activate agents.

Based on a variety of tasks to be solved in the implementation of situational management technology, multi-agent environment situational control is based on a hybrid architecture that includes deliberative (based on knowledge) and reactive (with the mechanisms of behavior such as "stimulus-reaction") agents [11].

Deliberative agents designed to perform analytical functions that result is situational management environment specification. Reactive agents implement process control management based on situational information (events) received from the environment or other agents in the implementation of communications. Thus emerging agents role groups (associations) with specialized functions within the group, both vertically (levels of hierarchy functions) and horizontally (the problem to be solved at this functional level).

Formation of agents knowledge about their environment is performed using monotonic and nonmonotonic logic mechanisms, which allows a more complete model of the world with regard to possible limitations and specific subject area, close to the picture of the world, formed in the imagination of man.

Agents autonomy allows to create similar agents in various locations of agent environment and to continue to make selection of the most effective agents for certain parameters in different models of interaction (competition or cooperation). Thus multiagent environment of situational management functioning permanence, flexibility and evolutionary is provided.

4 Conclusion

Development and support the functioning of SC associated with the creation of services runtime environment integration and use to meet the challenges of situational management.

1. Definition of services needed to solve specific problems of situational management is based on a formal specification of the problem in terms of the knowledge domain model.
2. Services search and integration realized by using agents with knowledge of the constituent elements of the situational management process.
3. Agent platform provides the tools for interaction-based service orchestration and choreography.

References

1. Michael Donley, Cornelius O'Leary, and John Montgomery: Inside The White House Situation Room: A National Nerve Center, https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/csi-studies/studies/97unclass/whithous.html
2. Kovalenko A.E.: Organization of agent-oriented situation management systems (in Ukrainian). In: 8th Science-Practical Conference on Decision Support Systems, pp.63-65. Institute of Mathematical Machines and Systems Problems NAS of Ukraine, Kyiv (2012).
3. OASIS Standards/ISO 15000, http://www.ebxml.org
4. ISO/IEC 11179, Information Technology – Metadata registries (MDR), http://metadata-standards.org/11179
5. Core Components Technical Specification – Part 8 of the ebXML Framework, ISO 15000-5: ebXML Core Components Technical Specification, Version 2.01, http://www.unece.org/fileadmin/DAM/cefact/ebxml/CCTS_V2-01_Final.pdf
6. Thomas Erl. Service-Oriented Architecture: Concepts, Technology & Design. Prentice Hall.
7. Wooldridge M.J. An Introduction to Multiagent Systems. - Chichester: John Wiley and Sons, Ltd., 2002. - p. 348
8. FIPA Agent Communication Language Specifications, http://www.fipa.org/repository/aclspecs.php3
9. FIPA Communicative Act Library Specification, http://www.fipa.org/specs/fipa00037/SC00037J.pdf
10. FIPA P2P NA WG6. Functional Architecture Specification Draft 0.12 - http://www.fipa.org/subgroups/P2PNA-WG-docs/P2PNA-Spec-Draft0.12.doc
11. Shoham Yoav, Leyton-Brown Kevin Multiagent systems: algorithmic, game-theoretic, and logical foundations. - New York : Cambridge University Press. (2009)

Comparative Review of the CRM Software's Solutions and Direction of Future Development

Pedja Milosavljevic, Milena Todorovic, Dragan Pavlovic

Faculty of Mechanical Engineering, University of Nis, Serbia
draganpavlovic10369@gmail.com

Abstract: In this paper is given the analysis and comparative review of the top best selling commercial Customer Relationship Management (CRM) software: SAP, Oracle, Microsoft Dynamics and NetSuite. This paper is intended to be used as a guide to help end-users in their vendor selection process. It is given the wide array of CRM vendors currently in the marketplace. In order to enable end-users to indentify which CRM software best meet their specific business requirements, they are categorized by functionality, customer focus, prizes and other criteria. This paper is primarily focused on the comparative analysis of the basic features of the CRM software. Also, it is presented future of the CRM software and possible direction of their further development.

Key words: CRM Software, Comparative Review, Cloud Computing, Software as a Service.

1. Introduction

Customer Relationship management is the strongest and the most efficient approach in maintaining and creating relationships with customers. Customer relationship management is not only pure business but also ideate strong personal bonding within people. Development of this type of bonding drives the business to new levels of success.

CRM significantly improves a company's internal organization in interaction with its customers and distributors, by establishing efficient communication mechanisms and accelerating the implementation of daily operative needs. In other words, the customer relationship management is organized in such a way that the needs, the demands and the expectations of customers are not only fulfilled, but also surpassed.

The CRM software represents an efficient way of business processes by which a company manages its contacts and the information about their customers, distributors, their respective needs and all the relevant information about market accesses.

The main goal of this paper is to be used as a guide to help end-users in their vendor selection process. It is also presented possibilities and directions of CRM future development.

2. Defining CRM

CRM represents systems, processes and procedures that help companies manage their customer relationships. CRM is a combination of processes, methods, systems, and technological know-how that will help the company deal with their customers effectively and efficiently.

The goal of the CRM is to retain and increase customers of the company. Customer relationship management is an approach designed specifically for corporations. It is a strategy whose main aim is to create as well as maintain long-term relationships with the corporation's customers. The main goal is to expand the company's client base.

CRM also presents a comprehensive system. It encompasses customer service but it goes beyond that. It is a system that involves and regulates all points of contact and business dealings with the company's customers. The CRM strategy is focused on how customer service representatives are going to deal with the customers. However, aside from the actual interaction, CRM will monitor and regulate all the communications used in the interaction.

For instance, CRM will have identified the reports that need to be generated after a customer has been given service. The CRM strategy will also be the basis for determining what notes will be written and filed about the customer interaction. As such, CRM will include how records on customers are kept. For instance, if a customer service representative makes a note about the customer's concern, the CRM software used may automatically compile the notes and update the general customer service database. Thus, when another customer service agent picks the call the next time the same customer calls back with the same concern, the customer will not have to explain his situation again. The CRM software ensures immediate access to information that would make the interaction between the customer and the customer service representative as smooth as possible.

A working CRM program would therefore include the acquisition of new customers and return business. It is concerned with isolating customers that would require a specialized line of service. A good CRM program and strategy will improve customer service. This can be done by developing a communication process that provides customer solutions in a timely manner. This can be done through the following ways:

- *Provision of information is essential.* Information about company's business products, products' uses, company's alternative services, troubleshooting guides, frequently asked questions and answers should be easily accessible and available. This can be provided on websites 24 hours a day and 7 days a week using a remote server.
- *Isolate customers' definition of quality*. It should be exerted every effort to ask every customer (through polls, surveys, etc.) about their idea of a winning customer service experience. With this on mind, company would be able to have a clearer picture and devise a strategy that would meet the requirements and expectations of its client base.
- *Fast and efficient customer support should be ensured.* After customers have made their purchase, company still has a responsibility to their customers. It should be devised a system that would allow to manage and schedule post-purchase client interaction post haste. This would enable managers of the company to evaluate dissatisfaction, common customer support concerns, repurchase times, probabilities and frequencies.

- *Install an efficient database for monitoring purposes.* Provide a structure that can allow keeping an eye on all interactions between the client and the company. All necessary sources and types should be included. It will also be beneficial if the system provides the same information that the customer sees and uses. This would eliminate or at least reduce confusion.
- *Be insightful.* Identification of an impending concern will provide a good image for the company by listening customer complains [1].

Customer relationship management is about managing all facts of company-customer relationships. This includes using all the advanced technologies, methodologies and systems that can ensure that company's customer will return and bring the economic business as well.

3. Criteria for Picking a CRM System for Business Needs

Which CRM system is right for an organization depends on several factors, and it is not a decision organizations can afford to take lightly. CRM is the path to customer loyalty, new business and increased revenues over the short and long-term, affecting every area of an organization's success. Value-added services can also be enhanced with CRM systems to establish stronger market presence.

Part of the CRM decision is determining which type of CRM system meets an organization's needs; organizations must evaluate whether a vertical, or a more general solution, will give them the best CRM outcomes. Increasing numbers of CRM solution vendors are streaming their mainstream CRM tools to seem more industry-specific; and likewise, industry-specific CRM vendors are striving to appeal to a broader range of industries, so the decision is becoming more complicated.

When considering *vertical versus a general CRM solution*, businesses should contemplate whether or not a vertical solution exists that truly meets their needs. This is what vertical solutions are designed to do, but not all can fully meet this functionality. While vertical CRM solutions may take longer to implement than more mainstream models, they can provide deeper levels of customizability. For example, features available in vertical CRM formats, like very specific requirements definition capabilities and code testing may not be available in a general CRM format. If an organization chooses a mainstream CRM solution when they need a vertical solution, they can find themselves engaging in layer after layer of testing before completing the evaluation process.

Company size is a key factor - In terms of scalability, organizations will maximize their opportunities for growth and meet their mission more effectively when a CRM solution matches their company size and plans for growth. Specifically, CRM systems must be ready to provide increases in performance during customer-heavy periods, like holiday seasons for retailers, and must be equipped to prevent lost resources during slower periods. CRM software should also allow customer agents to rapidly find customer records – even if this data comes from varying-sized batches of customer records.

A number of different industries suggest the "try before you buy" approach to products and this is an important step with CRM. If the person making the CRM selection is not versed in IT integration, the right questions may not be asked and the resulting implementation could be a disaster. Whether the company elects to implement hosted or on-premise solutions, most vendors can accommodate the try it first approaches. The cloud-based solution should be accessible through the Web browser and the full suite should be

made available. If all applications, functions and features cannot be tried before the contract is signed, it is wise to look for another vendor.

4. Comparative Analysis

The CRM market is currently experiencing rapid growth; analysts expect the global CRM market revenue to hit around at least $18 billion in 2012. CRM growth was driven by strong demand for marketing automation, sales automation and customer-service technologies [3]. There are many CRM vendors now and it is very important to choose the right one. Which CRM system is right for an organization can depend on several factors, and it isn't a decision organizations can afford to take lightly. These factors are:

- Industry Specific CRM or Mainstream Solutions;
- Company size;
- On Premise CRM or Hosted CRM;
- Software features (modules);
- Ease of use/implementation;
- Vendor evaluation [4].

According to [3] and [5] some of the best-selling CRM software are: Oracle, SAP, NetSuite and Microsoft Dynamics. A comparative analysis in this paper is done based on these results.

Microsoft Dynamics for customer relationship management empowers employees to boost sales, satisfaction, and service with automated CRM that's easy to use, customize, and maintain. Microsoft Dynamics business software offers a wide spectrum of affordable CRM solutions to help companies meet their specific needs. Microsoft CRM is unique in an increasingly crowded CRM software market to use the same code base for on-premise and Software-as-a-Service (SaaS) delivery models. It's also unique in its ability to support multiple public clouds for SaaS delivery.

The company's most recent release also allows cloud customers and business partners to install server-side code in the Microsoft cloud and data centers. This permits extensibility for a number of business processes not supported by most competitors.

The *Oracle CRM* includes: Oracle Siebel CRM, Oracle CRM On Demand, Oracle E-Business Suite, PeopleSoft Enterprise CRM and JD Edwards EnterpriseOne CRM. Oracle CRM has a reputation for high risk, and expensive deployments. In fact, failed implementations and frustrated users are the primary impetus for the rise and sky-rocketing growth of Cloud or SaaS CRM systems.

The company's hosted software, Oracle CRM On Demand, competes with Salesforce.com. CRM On Demand offers low-cost, hosted CRM solution with little or no up-front IT investment. With CRM On Demand, companies can accelerate sales, improve marketing and deliver consistently top-notch customer service.

NetSuite CRM is a hosted CRM application that provides sales reps with a 360 degree view of customers so companies can maximize customer revenue, from lead and opportunity management through order processing, customer service and support resolution, to renewal and up sell purchase management. NetSuite CRM also gives sales reps the ability to create quotes, place real orders, and gain-real time visibility into order status, overdue invoices, inventory, customer transaction history, up sell and cross sell management and commission compensation.

The *SAP CRM* software solution is a fully integrated CRM software that targets business software requirements of midsize and enterprise organizations across industries and market sectors. Similar to top competitor Oracle, SAP has more recently released two SaaS CRM products—SAP Sales on Demand which is a cloud CRM extension for the company's on-premise ERP software and SAP Business By Design which is a full ERP cloud suite that includes CRM.

SAP offers CRM on-demand solutions that are easy-to-use, Web-based, and available on a subscription basis.

The following Table 1 shows some of the basic advantages and disadvantages of CRM software [6].

Table 1 – CRM Product Properties

CRM Product Name	***Advantages***	***Disadvantages***
MS Dynamics CRM	- Good integration with Microsoft Office products - Reasonable sales force automation (SFA) - Strong technology foundation and architecture - Strong partner delivery network	- Heavy browser architecture -fat client - Weak marketing and customer base - Limited to small business organizations
Oracle on demand	- Integrates to Oracle Financials - Nice dashboard - Good data warehousing (lacks flexibility, but good presentation) - Strong sales force automation (SFA)	- Not as strong marketing automation or customer service - Lacks deep functionality offered by some other hosted vendors - Offline version is pretty bad - Allegedly poor customer service and turnover
NetSuite	- A single and enterprise wide integrated business software system - Role-based application navigation and security - A native thin-client solution designed for on-demand - Internet delivery - Several industry specific or vertical market solutions	-A difficult to use system - Troublesome customer support - High customer support costs - Poor company reputation - Difficult customization tools - Product upgrades are notoriously troublesome
SAP	- Backing by the largest application software vendor in the world - Isolated tenancy hosted delivery model is a welcome change from most other hosted CRM vendors - Strong ERP can be leveraged into the CRM	- The product is new, shallow and comparatively weak when compared to other hosted CRM vendors

The working area of CRM vendors is wide and varied. Some vendors are focused on the entire enterprise solution, while others have specific functional specialties such as Marketing Automation or Customer Service. In Table 2 CRM vendors are categorized by different criteria and presented in a matrix.

Table 2 – CRM Characteristics

Characteristics	**Vendor**			
	Microsoft Corporation	***Oracle***	***NetSuite***	***SAP***
Suite	x	x	x	x
Marketing	x	x	x	x
Customer Service	x	x	x	x
Call Center	x	x		x
Sales Force Automation	x	x	x	x
Hosted/SaaS	x	x	x	x
On-Premise	x	x		x
Enterprise	x	x	x	x
Mid-Size	x	x	x	x
Small	x	x	x	

As can be seen from the matrix Microsoft Dynamics and Oracle covers all area, while NetSuite doesn't have Call Center. On the other hand SAP is more focused on mid-size and big enterprises and doesn't have CRM software solution for small enterprises.

Each CRM software is consist of certain modules or features. The structure of the CRM software is usually consists of the following group of modules:

- *Customer Service;*
- *Sales;*
- *Marketing.*

Beside of these basic modules CRM software also contain other modules. In almost all CRM software there are differences in the names of certain modules, respectively modules can have exactly the same functionality but different names, or with the same name, but with completely different functions.

Table 3 shows a comparative review of the basic characteristics and modules of these software solutions [7], [8], [9] and [10].

Table 3 – Comparative Review of Basic CRM Characteristics

Vendor	Microsoft Corporation	Oracle	NetSuite	SAP
Product	• Microsoft Dynamics CRM	• Oracle CRM On Demand • Siebel CRM • Oracle E-Business Suite CRM • PeopleSoft Enterprise CRM • JD Edwards EnterpriseOne CRM	• NetSuite CRM + • NetSuite	• SAP CRM • SAP Business All-in-One • SAP Business by Design • SAP Business One
Modules	• Customer Service • Sales • Marketing • Go Mobile	• Sales force Automation •Partner Relationship Management •Marketing Automation • Customer Service and Support • Content Management •Innovation Management • Real Time and Historical Analytics • Desk Top and Mobile CRM • Data Model Customization • User Interface Customization •Enterprise Administration	• Salesforce Automation • Customer Service Management • Partner Relationship Management Marketing Automation • Mobile • CRM Analytics	• Marketing • Sales • Service • Partner Channel Management • Interaction Centre • Web Channel • Business Communications Management • Real-time Offer Management • Trade Promotion Management
Price in $	• 440 - 880 per user • 528 – 1761 per server	•Oracle CRM On Demand:70 user/month •Siebel CRM:70 user/month • Oracle E-Business Suite CRM: N/A • PeopleSoft Enterprise CRM: N/A • JD Edwards EnterpriseOne CRM: N/A	• NetSuite +: 129 /user/month • NetSuite: 499 month base fee and 99 /user/month	• 75 user/month

5. Future Development of CRM

With CRM technologies evolving, one of the important selling points has been the ability of CRM systems to provide a 360 degree view of the customer. Having an integrated view of the customer has helped organizations to keep all departments (Sales, Service, Finance etc.) on the same page and thus in turn provide a consistent service to the customer [11]. Essentially, what has been captured in the 360 degrees view is the transactional behavior of the customer.

The transactional data has been mainly classified under 3 sub-categories:

- Transaction data like sales orders, billing, receivables and warranties;
- Account data that would include receivables, outstanding balance, account hierarchies and credit history;
- Interaction data, essentially capturing communication from all customer touch points like emails, chat and call center.

Capture customer demographics through Social IDs – Social CRM. Social CRM, as such, is use of social media services, techniques and technology to enable organizations to engage with their customers. For this to happen, one of the first things that would need to be done is providing placeholders for LinkedIn IDs, Twitter IDs and Facebook IDs in the traditional CRM systems. Profile information and customer demographics can be collected by integrating the CRM systems with the social network using these identifiers. The collected information can be used to identify customer demographics, likes/dislikes, important milestones etc. Often customer demographics in the organization's database go outdated and there is a greater probability that the customer might have been updating this information more periodically in their social forums. A client manager with a mandate to interact closely with few select accounts would benefit if that extra information is made available about his or her customer beyond the typical transactional record. A lot of ecommerce websites have started using Facebook logins for customers. As per Facebook statistics, more than 80% of internet sites are integrated with Facebook. Once the customer logs in using a Facebook ID, you get direct access to the customer profile information and this data can be used to capture customer demographics, recommend products, suggest gift ideas for social connections etc [12]. They would still continue to exert significant importance and the behavioral information from the social media would complement to understand the customer better and forge more fruitful relationships. Recently, there has been a growing concern on the security aspects of personal information being shared across social media [13]. Integrating the traditional CRM systems with social media tools is very important else the CRM system could end up becoming a middleware solution storing only the transactional behavior of the customer.

Cloud CRM provides a simpler, faster, and more affordable way for businesses to take advantage of powerful technology tools that streamline and automate the way customer interactions are managed across touch-points [14], [15]. With Cloud CRM, all hardware and software components are purchased, installed, tested, and maintained by a third-party hosting provider at a remote site. The hosting service provider also stores and manages all customer-related data. Companies need nothing more than a standard Web browser to access and utilize the CRM application and its features. Cloud CRM makes customer relationship management easier and more cost-effective for businesses with IT and budget constraints. It provides all the standard functionality of licensed CRM solutions, as well as the security,

reliability and performance companies need to ensure smooth customer operations, without the time and expense associated with in-house systems.

As companies strive to remain competitive by operating in a more customer-centric manner, the need to keep staff members fully connected to client data at all times – even when they're on the road – has become increasingly important. Sales reps, field service workers, and other customer-facing personnel who travel frequently require any-time, anywhere access to the timely, accurate data that will enable them to acquire and support customers as efficiently and effectively as possible [16].

That's why leading analysts expect the market for *mobile CRM* applications to experience explosive growth in the near term. In fact, research firm Gartner predicts that industry expansion could be as high as 40 to 60 percent over the next two to three years. And, a Forrester Research study shows that nearly half of all businesses in North America and Europe have already deployed – or are planning to deploy – mobile sales force applications.

A *mobile CRM* application is a powerful, full-featured software solution that allows field workers – such as sales representatives, service staff, and support teams – to access and interacts with customer data while they're on the road [17], [18]. Using cell phones, Blackberry devices, Windows Pocket PCs, and other Web-enabled handheld appliances, users can retrieve and update customer-related information from back-end systems, manage opportunities and jobs, process orders, check inventory levels, and much more, just as easily as if they were in the office.

There are numerous benefits that can be achieved through the implementation and use of a mobile CRM solution. Among the greatest advantages are:

- Increased efficiency and productivity of field staff.
- Improved face-to-face interactions between clients and employees.
- Enhanced information flow and sharing throughout the entire organization, even among workers who are on the road.
- Increased accuracy and timeliness of information, through the ability of all field workers to make instant updates to data at any time.
- Reduced sales and service costs through remote automation, and the facilitation of faster, more informed decision making.

Many businesses are deploying *Software-as-a-Service* (SaaS) CRM. Software as a Service is a software distribution model in which applications are hosted by a vendor or service provider and made available to customers over a network, typically the Internet. SaaS is closely related to the ASP (application service provider) and on demand computing software delivery models. IDC identifies two slightly different delivery models for SaaS. The hosted application management (hosted AM) model is similar to ASP: a provider hosts commercially available software for customers and delivers it over the Web. In the software on demand model, the provider gives customers network-based access to a single copy of an application created specifically for SaaS distribution.

Benefits of the SaaS model include: easier administration, automatic updates and patch management, compatibility (all users will have the same version of software), easier collaboration, and global accessibility. In just a few years Customer Relationship Management has emerged as a powerful business trend. However, the best of the CRM is yet to come. By implementing a cloud phone system or managed social media support network, a

company can not only stay ahead of these constantly evolving trends, but also improve their customer loyalty and support today, rather than two or three years from now.

6. Conclusion

By taking the time to look at the key elements of growing business within any CRM system, every company can have better position for growth, enjoying the measurable benefits afforded in a robust system selected and implemented to fit the needs of the organization today and into the future. Based on all above mentioned in this paper it can be concluded that the choice of CRM software depends on the needs of the company and one of the key factor of successful implementation of CRM solutions is willingness to change, discipline in the system use as well as human and financial resources. The main reason for this is that this kind of project is characterized by a large scope of unseen costs and resources spent, which are difficult to predict.

References

[1] Gebert, H., Geib, M., Kolbe, L., Brenner, W.: Knowledge-enabled customer relationship management: integrating customer relationship management and knowledge management concepts, Journal of Knowledge Management, vol.7, no.5, pp.107 – 123, (2003).
[2] Taylor, S., Hunter, G.: The impact of loyalty with e-CRM software and e-services, Int. Journal of Service Industry Management, vol.13, no.5, pp.452 – 474, (2002).
[3] http://jck-consulting.com/who-has-the-top-crm-market-share-in-2012/
[4] http://www.crm-reviews.com/vendor-reviews/
[5] Business Software: Top 40 CRM software vendors' revealed 2011 edition, (2011).
[6] http://www.rishabhsoft.com/blog/a-crm-comparison-guide
[7] http://crm.dynamics.com/en-us/home
[8] http://www.oracle.com/us/solutions/crm/index.html
[9] http://www.netsuite.com/portal/products/crm/main.shtml
[10] http://www.sap.com/solutions/bp/customer-relationship-management/index.epx
[11] Gustafsson, A., Johnson, M., Roos, I.: The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention, Journal of Marketing, vol. 69, no. 4, pp. 210-218, (2005).
[12] King, S. F.: Citizens as customers: Exploring the future of CRM in UK local government, Government Information Quarterly, vol. 24, no.1, pp 47-63, (2007).
[13] Tanner, J. F., Ahearne, M., Leigh, T. W., Mason, C. H., Moncrief, W. C.: CRM in sales-intensive organizations: a review and future directions, Journal of Personal Selling and Sales Management, vol. 25, no. 2, pp 170-180, (2005).
[14] Cusumano, M.: Cloud computing and SaaS as new computing platforms, Communications of the ACM, vol. 53, no.4, pp 27-29, (2010).
[15] Zhang, Q., Cheng, L., Boutaba, R.: Cloud computing: state-of-art and research challenges, Journal of internet services and applications, vol. 1, no. 1, pp 7-18, (2010).
[16] Alahuhta, P., Helaakoski, H., Smirnov, A.: Adoption of mobile services in business – case study of mobile CRM, e-Business Engineering, Conf. Publ., pp. 531-534, (2005).
[17] Duran, R. E.: Extending CRM with mobile messaging: a case study, International Journal of Business Innovation and Research, vol. 4, no. 1, pp 15-29, (2010).
[18] Hartel, M., Bulander, R., Decker, M.: A literature survey on objectives and success factors of mobile CRM projects, Book of Proceedings: EURO mGOV, Second European Conference on Mobile Government, (2006).

Problems of the Small Passing Exams at Faculty of Mechanical and Electronic Engineering

Pedja Milosavljevic, Milena Todorovic, Miroslav Milovanovic, Dragan Pavlovic

Faculty of Mechanical Engineering, University of Nis, Serbia
milenatod1@yahoo.com

Abstract: In this paper is analyzed the number of the passing exams of the first and the second year at Faculty of Electronic and Mechanical Engineering University of Nis due to finding improvements to increase passing of the exams. One of the goals is modernization of the teaching process by introducing e-learning. In this paper are used some of the quality tools, such as: Pareto analysis, Statistical Process Control (SPC) and Ishikawa diagram. After defining defects (seven basic types of waste) that have negative influence on the final quality evaluation of higher education and how these defects can be eliminated, the Pareto analysis is done, which is used for establishing a vital minority of the exams that are critical for examination at both faculties. SPC analysis is performed on the exams that are classified as vital minority in Pareto analysis. Ishikawa diagram is applied, in order to find the causes that affect on small number of passed exams. The accent was placed on the comparative analysis of the passing exams at the Faculty of Mechanical and Electrical Engineering, as well as the improvements proposal. One of the major improvements is implementation of E-learning system.

Key words: Pareto analysis, SPC analysis, Ishikawa diagram, E-learning.

1. Introduction

Circumstances that are arising in the environment of Serbia, especially in the European Union, clearly show that the country needs a very thoughtful, organized and qualitative development of education because it is one of the main conditions for the development of Serbia to knowledge-based society capable of providing good employment of the population. These circumstances look for harmonization of education system in Serbia with the European educational system [1], [2].

The Bologna Declaration was signed in 1999, and with thus has been starting the Bologna process that aims to create a single European system of university teaching and research. Generally tends to create more flexible and efficient system of high education in Europe which would be competitive in the global market knowledge. The most important measures in the framework of the Bologna process are the introduction of ECTS, promoting the mobility of students and teachers and the adoption of a system of comparable degrees [2].

This paper presents the results of the study programs at the Faculty of Mechanical and Electrical Engineering at University of Nis, where has began the implementation of the Bologna process, where the data are used for student generations that are studying according to academic plan and program adapted to the Bologna Declaration.

2. Lean Six Sigma Method

Lean Six Sigma je a methodology that in recent years is the most mentioned in the literature and refers to procedures of improvements by the integration of methodology Lean and Six Sigma. Lean represents approach that look for improvement of flow values and the elimination of losses in a prompt manner. The goal of Six Sigma methodology is the complete elimination of defects in manufacturing and/or service processes [5]. When the flow is determined by the optimal value, Lean method focuses on the identification and designation of additional values together and then the Six Sigma tools are used to help understanding and reducing variation. The ultimate goal of Lean methodology is to eliminate waste. Waste is everything that adds cost or is time-consuming without adding value [6], [7]. Below is presented the seven basic types of waste identified in the process of education at Faculty of Electronic Engineering (table 1).

Table 1 – Seven basic types of waste

Waste	Definition	Elimination of waste
Defects	Non-passed exams, canceled exams, non-adapted education plan and program according Bologna process, lack of and/or inadequate literature, the lack of practical training, lack of students interest in certain school materials, oversized groups of students in lectures.	Continuous students surveys and the elimination of deficiencies in the teaching program, additional consultation, application of proven models of teaching, investment in laboratory equipment, stirring interest of students by following the worlds trends and new technologies, adjusted the number of students in groups so that the teaching is quality and student learning is more productive.
Inventory	Students who have not passed the exam, students who have given up of the exam, students who are rehearsing some subjects, the great variation in the number of exams passed per examination period, too extensive literature.	Compulsory attendance of students in lectures and exercises, fulfillment of all the pre-exam and exam commitments, additional working hours of teachers and professors with students in examination periods where the pass rate was less than optimum passage defined, adapting literature to the real needs of future engineers for quality learning activities and eject excesses parts from the literature.
Movements	The movement of students from one to another classroom, and the lack of information about the untimely date change, teaching, testing and pre-exam obligations.	Teaching in classrooms, updating information on the website provided by the faculty and bulletin boards about teaching schedule, pre-exam and exam obligations, as well as any announcements relevant to everyday student activities.
Waiting	Waiting for admission to next year, semester certification, exam application and other formalities in the performance of student services, the delay of students and professors in lectures and exercises, waiting for information about the exam and tests and consultation schedule.	On-line exams registration, extended time for student services work with students, control of student attendance, control of teachers and professors attendance and reporting identity card upon arrival or departure from class, daily updates on web sites of faculties and departments as well as on the bulletin board.
Transport	The transfer of information from student services to students, storage and collection of documentation in student services.	Bulletin board near the student services and in visible places optimized electronic database of student information and providing a simple and fast searching of printed stored documents.

Over-processing	Re-listening of lectures, the performance of pre-exam and exam commitments, inadequate quality of teaching, unclear and confusing literature.	Student motivation for better and quality work, review of the concept of lectures and exercises, eliminating and clarifying any uncertainties in the literature.
Over-production	Graduates of different professors whose profiles do not match the needs of the market and/or with the knowledge acquired at university is insufficient for job position and impractical in work with engineers from other profiles.	Cooperation between faculty and industry and monitoring needs of the company profiles for qualified personnel, constantly improving teaching programs harmonious international professional practice, engaging students in interdisciplinary projects which would lead to their better competence.

3. Pareto analysis

The aim of the Pareto method is the identification of vital minority of 20% of causes of the problems that need to focus resources in order to eliminate 80% of the problems and inconsistency arising in the process [8]. Software Pareto Analysis.Net v1.2, developed by company CIM Group is used to create the Pareto analysis.

Pareto analysis is based on the results of the exams progression of Faculty of Electronic Engineering at University of Nis for the academic year 2010/2011 that are studying according Bologna Declaration. In Pareto analysis is included the first and second year of undergraduate studies.

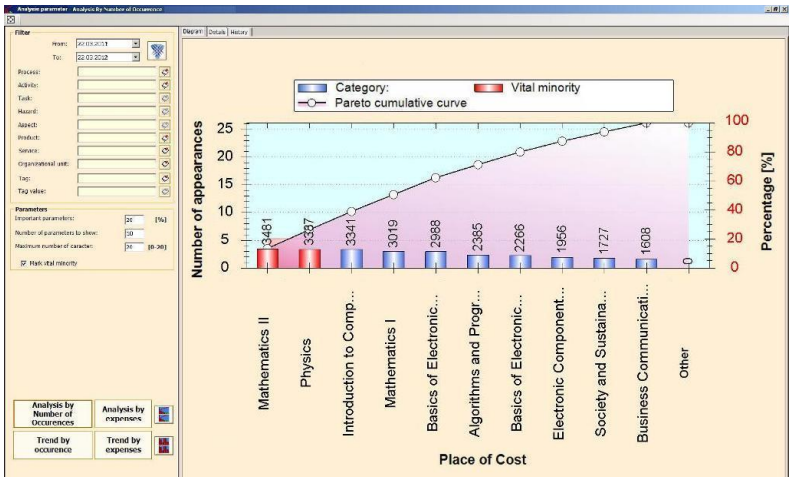


Figure 1 - Pareto analysis of the number of registered exams for the first year of undergraduate studies

It is presented the Pareto analysis (fig. 1) that indicates that a vital minority in the analysis of registered exams is consisted of two subjects: Mathematic II and Physics.

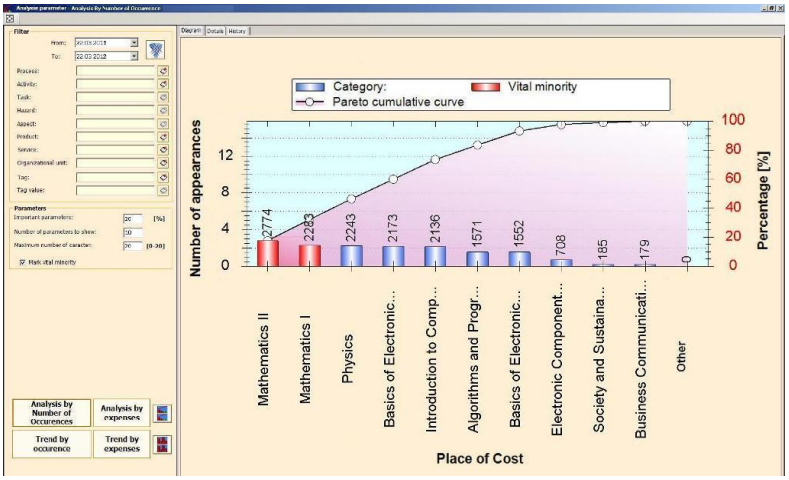


Figure 2 - Pareto analysis of non-passed exams for the first year of undergraduate studies

Pareto analysis (fig. 2) provides the analysis of non-passed exams for the same academic year, where vital minorities is consisted: Mathematic II and Mathematic I. The above mentioned exams represent critical exams due to passing of students, i.e. they present bottlenecks in continuous and successful studying, and on the same will be based further analysis in the paper.

4. SPC analysis

Statistical Process Control is an analytic tool for decision making that allows it to see when the process works properly and when not [9]. After Pareto analysis and determination of vital minority of subjects which are critical for passing the exams, it is applied SPC analysis. Control points are examination periods and depending on subjects their number is different. This is because the semesters are different in which subjects are listened. The average value, the upper and lower specification limits are different for each subject and depend on the number of exams passed, expressed as a percentage. Software SPC.Net developed by company CIM Group is used in SPC analysis. SPC graphs are presented for the first year of undergraduate studies that represent a vital minority (fig. 3 and 4).

According to control charts of the observed subjects can be noticed that the process can vary widely and that is outside the specific limits that are curtained by expectations of the Bologna process. In the figures of the capability of subject can be traced defined specification limits, as well as limits that software determines according to the entered values. According to the values C_{PK} can be observed that the process if far below the level of Six Sigma parameters.

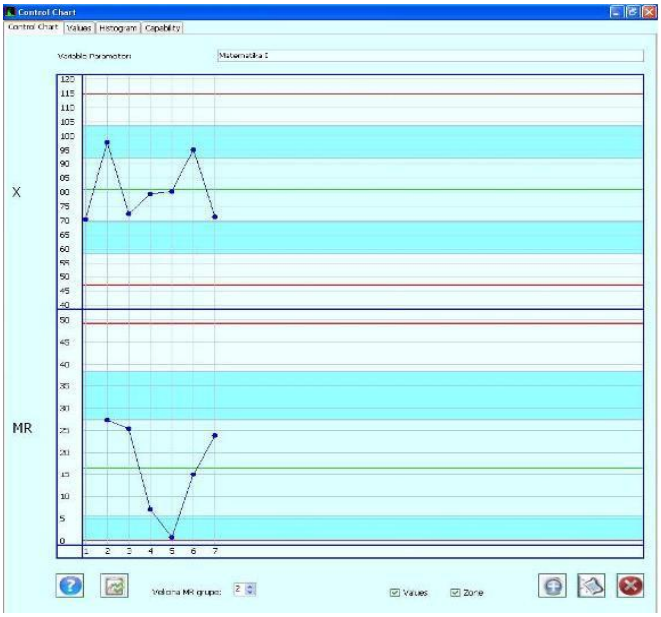


Figure 3 – Control chart for subject Mathematic I with control points: January, April, May, June, September, October I and October II

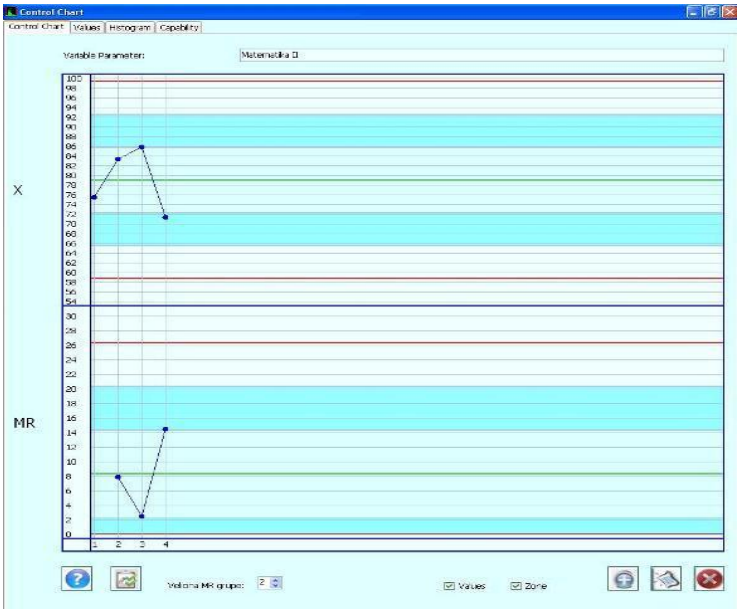


Figure 4 – Control chart for subject Mathematic II with control points: June, September, October I and October II

5. Comparative analysis of the results at the Faculty of Electronic and Mechanical Engineering

The goal of comparative analysis of the passing exams of the first and the second year of undergraduate studies at Faculty of Electronic and Mechanical Engineering, for the academic year 2010/2011, is multifaceted. Since these are technical faculties whose teaching programs overlaps to a certain extent, some subjects which are hold at both faculties are comparable as well as the result of number of passed exams and it can be determine whether there is a common problem for these subjects. The problematic subject can primarily open the issue of quality of teaching program, then the issue of student motivation to learn and finally whether and to what extent the provided material is necessary to students. Also, a comparison can show if the results for the same subject are better at one faculty then to see the results on the other faculty. In that case, it should be considered is it possible to increase the number of passed exams by applying a successful model of teaching from one faculty to other. Analysis of passed exams per examination periods can answer which examination period is most productive.

Below it is shown a comparative analysis of the non-passed exams at both faculties for exams in the first year for January examination period (fig. 5 and 6).

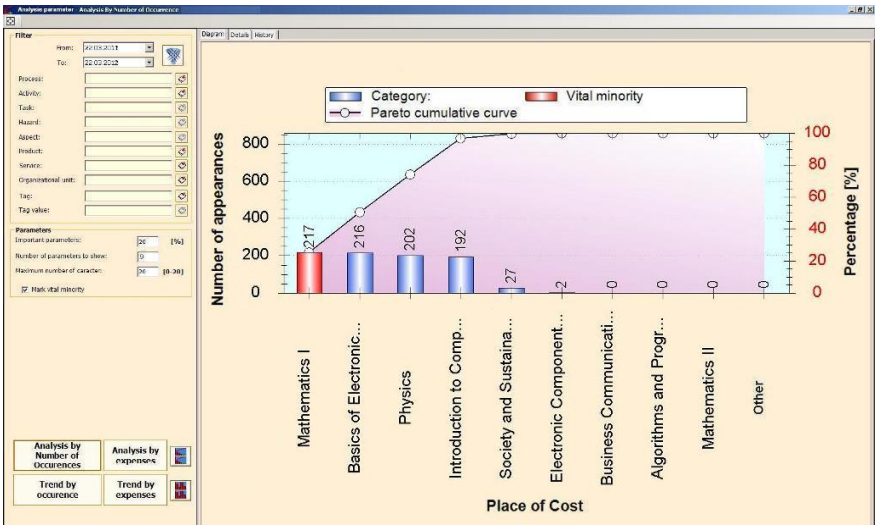


Figure 5 - Pareto analysis of non-passed exams for the first year of undergraduate studies in January examination period – Faculty of Electronic Engineering

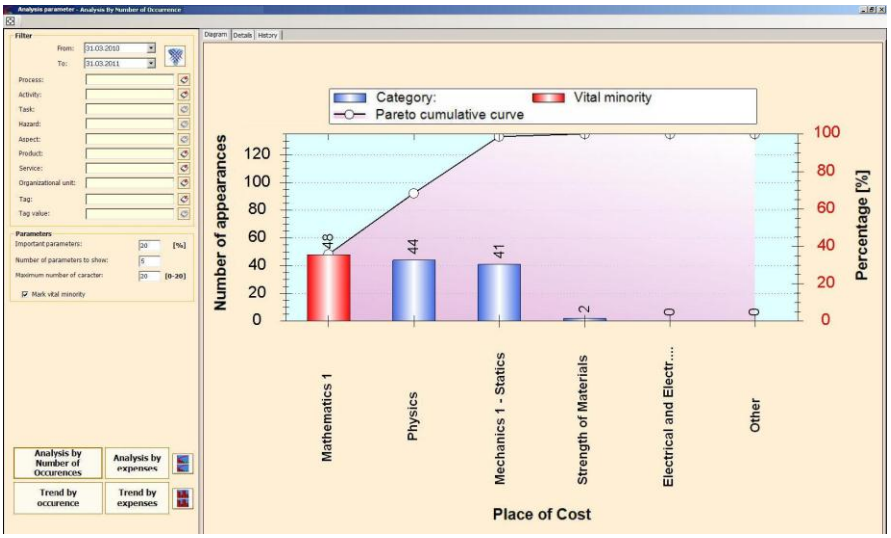


Figure 6 - Pareto analysis of non-passed exams for the first year of undergraduate studies in January examination period – Faculty of Mechanical Engineering

It can be concluded that vital minority of subjects, which are critical for passing the exams at both faculties, is Mathematics I. General subjects at both faculties, such as: Mathematics, Physics, Mechanics and Electrical Engineering, represent a bottlenecks not only in the first year, but represent non-passed exams during further studying. One of the reasons is that on the first year are students with different previous knowledge and skills to fit into the new environment.

This problem should be the focus of both faculties and the management and teaching staff in order to enable the success of these subjects according to the Bologna Declaration. These subjects represent basic techniques and without them further upgrading of knowledge would be inefficient. In the next chapter are given the improvements in order to eliminate this problem.

Below is a comparative analysis of SPC control charts for the subject Mathematics I for both faculties, which shows the variation of the number of passed exams through the examination periods (7 and 8).

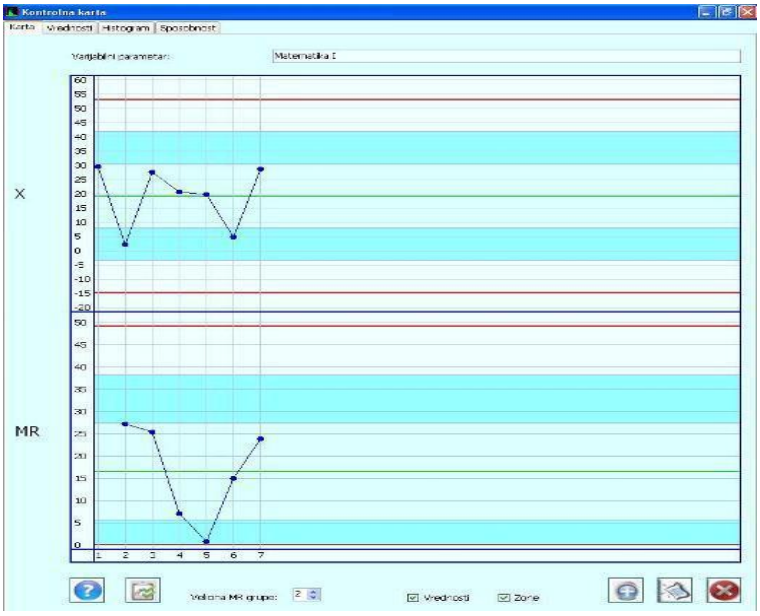


Figure 7 – Control chart for subject Mathematic I with control points: January, April, May, June, September, October I and October II – Faculty of Electronic Engineering

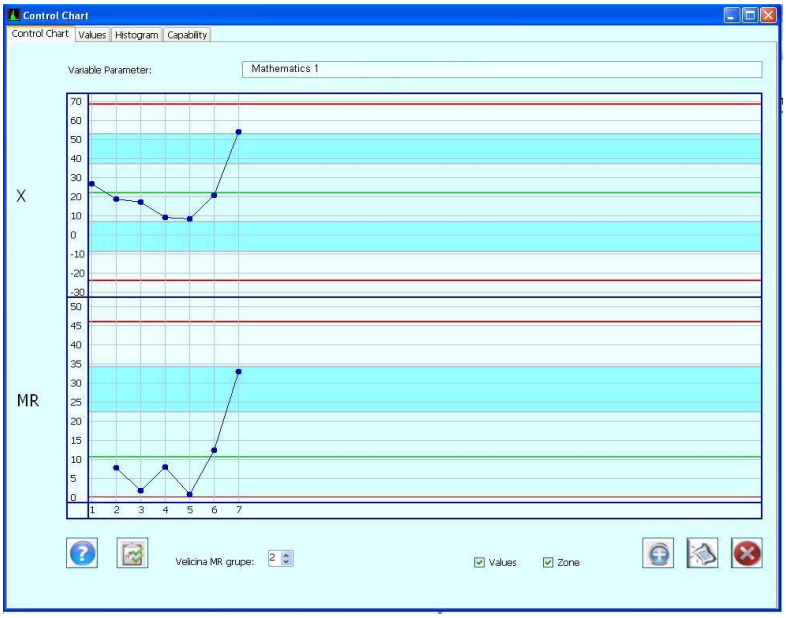


Figure 8 – Control chart for subject Mathematic I with control points: January, April, June, September, October I, October II and October III – Faculty of Mechanical Engineering

According to the comparative control charts for the Mathematics I at two technical faculties, it can be noted that the process varies widely and it is outside the specification limits that are specified with Bologna process. Interestingly, it can be noted that in both cases the graphics are about the same shape, i.e., the number of passed exams is growing especially in the period from September to October, when most students pass exams for enrollment into the next year.

6. Improvements

6.1 Generally improvements

What can be concluded already in the previous part of this paper is that the improvement in the teaching process and increasing success of students is a complex problem whose solution depends on a number of parameters and required activities. First of all it is necessary to continually survey the students and defining the problems. For subjects, where is small number of passed exams, it is necessary to provide additional consultations, to adjust literature to the real needs of future engineers and eject excess materials from the literature. Also, if necessary, adjust the number of students in groups at lectures and exercises. No matter how successful passing the exams is, it is necessary that the teacher constantly strives to arouse interest of students by following the world trends as well as applying of new technology in the teaching process as much as possible. The Bologna process implies also the mobility of teachers in order to introduce them with methods of teaching at other Universities and if they recognize a better method to implement it at home faculty. For examination periods in which it is determined that the passing number is less than optimal, it is desirable to provide additional working hours with students.

Student services also play an important role in this process and it is necessary to improve the software systems in order to update all data required for the timely informing students. It is preferable to develop the web portal with all information, in order to improve informing of all interested groups. Also it can be noted is the lack of practical training of students so it is necessary to intensify the cooperation between faculties and industry, tracking needs of the companies for qualified personnel, constant improvement of the teaching programs according with market needs and investing in programs of national and international internship training. Also, it was shown that the involvement of students in interdisciplinary projects increases their level of competence and opens the possibility to share experience with students from other educational backgrounds.

6.2 E-Learning

One of the major suggested improvements is implementation of E-learning system. The environment of higher education is evolving. Rising costs, shrinking budgets, and an increasing need for distance education are causing educational institutions to reexamine the way that education is delivered [10]. In response to this changing environment, e-learning is being implemented more and more frequently in higher education, creating new and exciting opportunities for both educational institutions and students.

E-learning is the expression broadly used to describe “instructional content or learning experience delivered or enabled by electronic technologies” [11]. The broader definition can include the use of the Internet, intranets/extranets, audio and videotape, satellite broadcast, interactive TV, and CD-ROM, not only for content delivery, but also for

interaction among participants [12]. More recently, this definition can be further expanded to include mobile and wireless learning applications [13], [14]. The main reasons for implementing e-learning are [15]:

- Increasing the availability as well as lowering the costs of information and communication technologies.
- Capacity of ICT to support and enrich traditional educational practices through learning based on sources and non-synchronized communication.
- The need for flexible access to learning opportunities from distant places such as homes, workplaces and conventional educational institution.
- Requirements of isolated and independent students for alternative approach to education services.
- Increasing the competitiveness in the growing education market.
- The need among educational institutions to keep pace with the times (with integration of ICT) to attract the attention of parents, students and other fund donors.
- The belief and expectation that on-line learning will reduce costs and increase productivity and efficiency of educational institutions.

The main obstacles to the implementation of e-learning in educational institutions are:

- Motivating teachers to integrate ICT into their teaching;
- Response of users (students);
- The problem of financial planning ICT of resources for e-learning.

The latest trend in the field of e-learning is the use of mobile learning. Increasing the number of mobile devices and Internet network expansion has created conditions for an increasingly common utilization of mobile learning.

Mobile learning, or m-learning, has been defined as learning that takes place via such wireless devices as mobile phones, personal digital assistants (PDAs), or laptop computers.

In other words, any sort of learning that happens when the learner is not at fixed, predetermined location, or learning that happens when the learner takes advantage of the learning opportunities offered by mobile technologies, is consider as m-learning [16].

7. Conclusion

Development of higher education in general, particularly bachelor and master studies, is facing a number of challenges because these studies are of the most importance for the development of Serbia. According to the Strategy for the Development of Education in Serbia until 2020, it is recommended using the methodology and technology of e-learning as a supplement to traditional learning.

In order to maximize this potential, e-learning implementations should bring to satisfy the needs and concerns of all stakeholder groups as much as possible. Educational institutions should integrate technology into classrooms to facilitate lecture delivery and create new technology mediated learning opportunities for students. In this paper, based on locating and defining the problems in the reform of higher education at two technical faculties, are suggested the measures and procedures to eliminate any deficiencies that may exist in the reform process. The methods that can affect on improvement of the quality of teaching process and student success in the exam periods are defined. It can be expected that the implementation of these improvements will increase the number of passed exams while maintaining the quality of the teaching. Once again it should be emphasized the importance

of continuous monitoring of parameters that affect on students passing the exams, the quality of teaching and in accordance with the results to perform appropriate procedures for improvement of these parameters.

References

[1] The Government of the Republic of Serbia, Ministry of Education and Science: Education Development Strategy in Serbia until 2020, Belgrade, (2011).
[2] Mijic, Lj., Lovcevic, B., Lazarev, S., Stanojevic, D.: Evaluation of the pedagogical work of teachers and associates in order to ensure the quality of teaching, Book of proceedings 37th National Conference on Quality, Kragujevac, Serbia, (2010).
[3] House of Commons Education and skills Committee: The Bologna Process, Fourth Report of Session 2006-07, London (2007).
[4] Quinn, A., Lemay, G., Larsen, P., Johnson, D.: Service quality in higher education, Total Quality Management, vol.20, no.2, pp. 139-152, (2009).
[5] Pryor, M. G., Alexander, C., Taneja, S., Tirumalasetty, S., Chadalavada, D.: The application of Six Sigma methodologies to university process: The use of student teams, Journal of Case Studies in Accreditation and Assessment, vol.2, (2010).
[6] Mladenovic, M., Milosavljevic, P.: The road towards a Lean Six Sigma company, International Journal Total Quality Management & Excellence, vol.38, no.3, (2010).
[7] Stoiljkovic, V., Stoiljkovic, P., Stoiljkovic, B.: Implementation Lean Six Sigma concept in manufacturing and service organization, International Journal Total Quality Management & Excellence, vol.37, no.1-2, (2009).
[8] Pavlovic, D., Milosavljevic, P., Mladenovic, S.: Application of Lean Six Sigma Method in Education process, Book of Proceedings: Total Quality Management -advanced and intelligent approaches, with 2nd Special Conf. Manufuture in Serbia, Belgrade, (2011).
[9] Stoiljkovic, V., Milosavljevic, P., Randjelovic, S.: Six Sigma Concept within Banking System, African Journal of Business Management, vol.4, no.8, (2010).
[10] Wagner, N., Hassanein, K., Head, M.: Who is responsible for E-Learning Success in Higher Education? A Stakeholders' Analysis, Educational Technology & Society, vol.11, no.3, pp.26-36, (2008).
[11] Ong, C. S., Lai, J. Y., & Wang, Y. S.: Factors affecting engineers acceptance of asynchronous e-learning systems in high-tech companies, Information & Management, vol.41, no.6, pp.795-804, (2004).
[12] Industry Canada: The E-learning E-volution in Colleges and Universities: A Pan-Canadian Challenge, from http://www.cmec.ca/postsec/evolution.en.pdf, (2001).
[13] Kinshuk, Suhonen, J., Sutinen, E., Goh, T.: Mobile Technologies in Support of Distance Learning, Asian Journal of Distance Education, vol.1, no.1, pp.60-68, (2003).
[14] Lehner, F., Nosekabel, H., Lehmann, H.: Wireless eLearning and Communication Environment, e-Services Journal, vol.2, pp.23-41, (2003).
[15] Faculty of Transport and Traffic Engineering, University of Belgrade: E-Learning in Theory and Practice, Manual for easy creation of online courses, **(**2007).
[16] O'Malley, C., Vavoula, G., Glew, J.P., Taylor, J., Sharples, M., Lefrere, P.: Mobilearn WP 4 – Guidelines for Learning/Teaching/Tutoring In A Mobile Environment, (2003).

Facebook and Twitter a platform to engage in a positive learning

Saraswathy Thurairaj, Swagata Sinha Roy, Kavitha Subaramaniam

Universiti Tunku Abdul Rahman(UTAR), Selangor Malaysia

tsaraswathy@utar.edu.my, swagata@utar.edu.my, kavitha@utar.edu.my

Abstract. Learning is a multifaceted human endeavor which consists of three main components; learning outcomes, mental processes and interactive processes between learners and their social environments. Therefore an interesting platform to conduct learning activities will encourage students' participation positively. According to Kabilan, Ahmad & Abiding (2010) social networking sites such as Facebook is one of the latest examples of communication technologies that have been widely-adopted by students, and thus it has the potential to become a valuable resource to support their educational communication and collaboration with their respective faculties. This paper explores the significance of ICT in education especially in a language classroom. Activities carried out were aimed at drawing the learners' usage of Facebook and Twitter in acquiring knowledge and learning in a language classroom. The methodology employed was an experimental study using the social networking sites in a language classroom which used questionnaire and in-depth interviews in collecting the data. Since the data were on nominal scale, the mean and mode of the Statistical Package for Social Science (SPSS) methods were used to compute and analyze the stated hypotheses. In addition this research has successfully identified the advantages of employing these new platforms in education and its potential issues for the betterment of student learning.

Key words: Learning, Facebook, Twitter and Social network.

Introduction

Learning is multi-dimensional and in today's highly advanced technological environment there are more and more tools to assist general learning. The buzzword today being 'lifelong learning', learners at all levels and of all ages must constantly adapt to changes in the learning sphere. One major learning tool is the internet, which has expanded by providing many powerful features and with such expansion; people have found that the internet can be used to connect people instead of only seeking information giving rise to social networking. The statistics for social networking users are startling the world over and nationally. A relatively high percentage, 72% of total internet users in Malaysia and from the comScore World Metric, found the social networking penetration in Malaysia was 66.6% in December 2008. Research has also shown Facebook as the top social network followed by Twitter which ranks as the 8th popular social network in Malaysia. In addition, the largest age group of Facebook and Twitter users is seen as ranging from 18-24 consisting mostly of teenagers and/or students. To have either of these accounts is free, easy and simple, thereby

attracting teenagers and/or students. They tend to use Facebook and Twitter as a means of gathering information, and for social networking with their friends through information sharing.

Background

The development of social networking websites such as Facebook and Twitter are becoming more popular among young adults especially the college and university students. These social networking websites allow the students to share information, conduct online discussions as well as have a direct live communication with their friends, family members or even lecturers. Such attributes may facilitate collaborative reading and writing (Warschauer and Kern, 2000). A lot of research findings have discussed the advantages of using web technology such as blogs and Wikis. However there are not many studies done on Facebook and Twitter to see whether these sites bring positive impact on English language learning among college and university students. This is because English is an official second language in Malaysia and it is widely used at every educational level.

Statement of the Problem

The development of social networking websites such as Facebook and Twitter are becoming a cliché among young adults, a platform for them to share information. However, there are not many studies done on these websites to see whether they will bring positive impact on English language among college and university communities. Thus, this research aims to explore the potential of using Facebook and Twitter as an educational platform other than in the classroom for positive language learning. This study is crucial to serve as a guideline for the education industry to investigate whether social networking sites will interfere or facilitate language learning among social network users through the understanding of the main impact of social networking sites on the language learning process of users.

Objectives

The primary objective of conducting this research is to determine whether Facebook and Twitter will interfere or facilitate language learning among both social network users through the understanding of the main impact of social networking sites on language learning process of users. Next, this research also helps to determine whether Facebook and Twitter will help users of both social networking sites in improving communication and writing skills through the activities and features that come with the sites.

Finally, the educational impact of Facebook and Twitter on users will be learned and determined throughout this research. In other words, this research will help researchers to identify whether Facebook and Twitter will provide the environment that will enhance users' proficiency in language or distort their language learning process.

Research questions

1. Do Facebook and Twitter lead to English language learning?
2. Do Facebook and Twitter improve communication and writing skills in English?
3. Do Facebook do Twitter really provide an environment for language learning?
4. How do Facebook and Twitter benefit UTAR students from the aspect of language?

Significance of Study

This study is important to show that social networking sites such as Facebook and Twitter can be one of the potential tools in encouraging language learning due to its many advantages. Furthermore, this study will also prove that English language writing and communication skills can be developed from the use of Facebook and Twitter. This research will identify the advantages of using Facebook and Twitter as part of the language learning tools and that these sites can be a source of guidelines for educators to adopt in teaching students English language and literature.

Review of Sources

Williams and Chinn (2009) states that information technology literacy skills can develop in an effective way in interdisciplinary settings and that by implementing experiential learning activities, particularly the use of Web 2.0 technologies, it can improve the involvement and engagement of "net generation" students in learning various disciplines. The research was related to a sports management course. The main aim was to maximize student involvement by letting students explore, reflect critically and engage in active feedback. The use of web technologies such as Facebook sparked interest in the students, leading to high energy and excitement in the classroom. Facebook helped in communication with peers as well as in placing advertisements by location. Even in the evaluation of assignments, the teaching fraternity's support of using web technologies was seen. The students had to research Web 2.0 tools and analyze the value of the tools in tandem with the requirements of the assignments where individual and group grades were assigned accordingly.

Tucker and Courts (2010) postulates that learning, 'sharing and ownership of knowledge' is changing due to the widely diverse technologies available in higher institutions of learning. Considered a 'dazzling invention' (Harkin, 2009), the net savvy population today find the internet indispensable. Tucker and Courts found that students who constantly turn to the net for information are also exposed to multimedia tools that facilitate and enrich their learning experience within the four walls of the classroom. There are both audio and video multimedia resources available and using audio resources helped in gaining the interest and increasing participation of the students in the classroom. Subjects considered to be previously boring became interesting overnight as instructors began to make use of power point slides and other multimedia tools in the classroom. Although not a new concept, the task of integrating technology into the classroom gives educators cutting edge methods which make the fixed curriculum more acceptable to present day learners.

The internet has created and encouraged new learning environments. Its evolution from web 1.0 in the 1990s into Web 2.0 today highlights much collaboration and interactivity (Enonbun, 2010). Enonbun posits that learning is an active process where the learner, through the use of the five senses, look at meaning out of the information being passed around. Being a primarily social activity, a chain of people such as teachers, students, peers and so on are involved. Further in the discussion, Enonbun reports that Web 2.0 technologies, namely blogs, wikis, social networks are now a major student occupation, thus demanding for instructors to constantly upgrade their technical skills and use them in innovative ways.

Information now can be accessed almost instantly and students can learn at their own pace as they solely control the direction and progress of their own learning. Enonbun further elaborates that with the use of wikis, blogs, podcasts, social network and virtual spaces, along with encouragement of instructors, assist in ensuring that learners from various ethnic groups (due to the heterogeneous composition of the student population under study) have a most conducive learning environment. On the other hand, Web 2.0 can limit resources which are computer based, cast doubts on the integrity of the said work or materials, give rise to plagiarism due to easy availability of various materials as well as increase the lack of privacy as anyone (outside of the students' (professors and peers) can see and evaluate what the students have written or uploaded.

The SNAP (Social Networking for Academic Purposes) Platform developed at Victoria University in Melbourne, Australia looks at the evolution of the social web which encourages the formation and development of e-learning platforms that actively engage learning communities. Kirkwood (2010) addresses a need for students to find a platform for locating one another via learning groups with similar interests and needs. Although most students spend too much time on Facebook for social networking purposes, they still need avenues to engage in networking for academic reasons. Kirkwood believes that Web2.0 technologies and services have made online peer assisted learning a possibility. Learners share their ideas and resources where they are expected to cooperate with one another thereby encouraging the social aspect of learning where the learning is learner-driven. In the personal learning environment, the instructor becomes a knowledge creator, mentor, facilitator all rolled into one, allowing students to pursue their own learning strategies and techniques, and that too, at their own pace.

More and more tertiary educators are encouraging open-ended, learner-centric, flexible learning platforms, where students can utilize the many varied and innovative web learning tools. In a research article by Hanson and Thomas (2009), the researchers are concerned with how computers can transform the way people learn language using "socially oriented" software. SNSs (Social Networking Sites) began to appeal to the younger generation with the introduction of Friendster, MySpace and Facebook. In recent times though, people have fallen out with Friendster and MySpace due to various social problems. Facebook continues to be a hot favourite among various SNSs. Having said that, more and more online communities that have sprouted over the last two decades are changing the ways of communication among the populace. One popular concept is the 'café' styled concept where the ambience of a real world café is recreated to encourage the establishment of a global community of language learners. The boundaries between study and leisure, teacher and learner start blurring, containing online communities of like-minded language learners. The buddy-tutor roles can be played out in such communities, increasing the ease of learning.

The various sources reveal a great tendency of looking at web tools and internet based learning as not only favorable but impactful and desired.

Methodology

Questionnaires were designed to match the objectives and the research questions for this research project. In collecting data for this research, the materials that are used include demographic and background information on the usage of Facebook and Twitter during the students' learning and activities outside their classroom. It is followed by a series of questions to find out users' activities on Facebook and Twitter, the features that the users use most often. Besides, the researchers would also like to explore the perceptions of the users on these social networking sites in helping to improve the users' English. For qualitative approach, in depth-interviews with the students were conducted. Target participants were interviewed because they have experienced using Facebook and Twitter in the ESP classroom as a positive learning platform and vice versa. Thus, these in-depth interviews based on the views and experiences of the students in the task have been more insightful. These in-depth interviews allowed more meaningful follow–up questions to be asked and answered and resulted in more extensive findings. The collected data was analysed using the SPSS mean and mode statistical approach.

The population for this research were all the year 3 semester 3, Accountancy and Global Economics Studies students. These students' age range is between 20 and 24. A total of 143 students from a private university within the Kajang District of Selangor, University Tunku Abdul Rahman (UTAR) was selected to undergo this study in order to obtain noteworthy results. The Accountancy and Global Economics Studies students were chosen because they have some exposure to the interaction via Face book and Twitter especially for their Report writing class, whereby allowing the researchers to get authentic data from the students. All the participants had passed their first year Business English and second year English for Management subjects before pursuing Report Writing in their final year. Therefore in terms of writing ability in English, these students were rated as good. Self-administered questionnaires were considered for gathering data at the empirical level. A total of 35 questions was administered to 143 students (84 females and 59 males) at UTAR, 58% of the participants of whom were from the Accountancy course and 42% were from Global Economic Studies. Information on the subject group is demonstrated in Table 1.

TABLE 1: INFORMATION ON THE SUBJECT GROUP

Variables	**Number (N)**	**Percentage (%)**
Gender		
Female	84	59
Male	59	41
Total	143	100
Courses		
Accountancy	83	58
Global Economics Studies	60	42
Total	**143**	**100**

Findings

TABLE 2: FACEBOOK AND TWITTER LEAD TOWARDS LANGUAGE LEARNING

Number (N)	Features	Percentage, %
1	Surfed Facebook and Twitter very often	80%
2	Language setting used (English)	95%
3	Facebook's and Twitter's spellcheck able to improve user's grammar	40%
4	Facebook and Twitter are able to enhance their language learning	45%
5	Facebook and Twitter are helpful	78%
6	Facebook and Twitter help their diversity of vocabulary	50%
7	Facebook and Twitter have strengthened English language learning in many ways	60%

Table 2 above shows that 80% of students surfed Facebook and Twitter very often. In terms of language setting used, 95% of them preferred to use English settings in their Facebook and Twitter accounts. In addition, "spellcheck" facility in Facebook and Twitter had improved users' grammar holistically with a total 40% positive response. This led to 45% respondents stating that Facebook and Twitter are able to enhance their language learning. From the findings, 78% of the respondents also claimed that Facebook and Twitter are helpful in the sense that both social networks are able to serve as a platform for them to explore English language learning. The usage of English on Facebook and Twitter also helped to expand their vocabulary range. 60% stated that their English language skills were strengthened through learning vocabulary from their friends, reading the articles posted by friends, and participating in English language learning games. In short Facebook and Twitter do lead to positive language learning.

TABLE 3: A SUMMARY OF PERCENTAGE SCORE OF THE ACTIVITIES USERS HAVE CONDUCTED ON FACEBOOK AND TWITTER

Number (N)	Features	Percentage %
1	Have updated their status or posted comments occasionally	45%
2	Have pointed out grammatical errors made by their friends or other users a few times	60%
3	Have supported that by using Facebook and Twitter, it will enhance users' communication and writing skills in English	55%
4	Have checked their spelling before posting or tweeting	55%
5	Have referred to dictionary to check the meaning occasionally	35%
6	Have used abbreviated words but did not apply the same practice in their other work	94%

Based on Table 3 above, 45% of the respondents had chosen to update their status or posted comments occasionally on Facebook and Twitter. It suggests that the majority of UTAR students are likely to publish their statuses or comments in English. In relation to language learning, 60% declared that they had pointed out grammatical errors made by their friends or other users a few times. The result shows that the majority of UTAR students are helpful in promoting the correct way to use English to users who had posted statuses or comments. Furthermore, 55% of the respondents supported that using Facebook and Twitter will enhance users' communication and writing skills in English because 55% of the respondents check their spelling before posting or tweeting. Thus, most of the respondents are concerned about their spelling accuracy. However only 35% has referred to the dictionary to check the meaning occasionally. Although 80% of the respondents usually used abbreviated words, among them, 94% did not apply the same practice in their other work.

The respondents reported that they rarely use abbreviated words in assignments and exam scripts. This shows that the users are able to switch their mind and apply different sets of words to cater to their purpose and to the occasion. Thus, most of the students are able to communicate in proper language despite the frequent use of abbreviations in social network.

TABLE 4: SUMMARY OF THE MEAN SCORES OF THE PREFERRED ACTIVITIES ON FACEBOOK AND TWITTER

Number (N)	Features	Ranking (Mode)	Ranking (Mean)
1	Users prefer to check messages, tweets or notifications when they are surfing the site	1	2.05
2	Users prefer to check their friends' activities or information	2	3.4
3	Users prefer to view photos or videos posted by friends	3	3.9
4	Users prefer to chat	4	4
5	Users prefer to update personal statues	5	4.55
6	Users prefer to share articles, videos, news, URL links etc	6	5.5
7	Users prefer to play games	7	6.15
8	Users prefer to send emails	8	6.4

Table 4 above shows that users prefer to check messages, tweets or notifications when they are surfing the site. This showed up as the lowest mean score of 2.05. Next, they prefer to check their friends' activities or information which showed the mean score of 3.4, followed by viewing photos or videos posted by friends with a mean score of 3.9 to keep them updated with their friends' most recent activities. Thus, the top three general uses of Facebook and Twitter are checking notification, friends' activities and viewing photos and video.

Discussion

It has been noted that students agreed that Facebook and Twitter should be incorporated as learning tools for language learning as shown in Table 2. Facebook and Twitter do lead to positive language learning. Based on the interviews conducted, students stated that activities through social media are very interesting and it makes learning easier besides helping them improve their communication. Thus, making Facebook and Twitter as learning tools is claimed to be fun and it is welcomed by the students.

It is proven that Facebook and Twitter are used widely to communicate in English as shown in Table 3. There are various activities preferred by the students and most importantly everything leads to positive language learning. Thus the students do improve tremendously in their language as well as their writing skills.

The students also claimed that they do learn and improve their language from the preferred activities on Facebook and Twitter as shown in Table 4. It is in fact a learning process when they check messages, tweets or notifications while surfing the sites. Besides,

the students also confirmed that they learn new sets of vocabulary when they check their friends' activities or updates.

Findings showed results that meet these research objectives. This is because students are very open to the usage of Facebook and other social networks to support classroom activities. With the features on Facebook and Twitter, they can enhance classroom discussion and make it more interesting. According to Thurairaj and Roy (2012), besides keeping in mind that the teaching materials should help their students to master the language, teachers also have to make sure that the teaching materials are interesting. Thus activities trough Facebook and Twitter will definitely be more interesting in nature to enhance positive language learning.

Conclusion

This research has shown positive results of using Facebook and Twitter as a platform to engage in English language learning. Besides, it fulfills all the objectives where Facebook and Twitter can lead to English language learning because most of the students think that it will be able to diversify their vocabulary. Besides, it was found that Facebook and Twitter improved the communication and writing skills in English as most of the respondents use the spell check program for errors. Since most of the respondents are very open to the usage of Facebook and Twitter to support classroom activities, Facebook and Twitter should be used extensively in the classroom. That way it can provide an environment for language learning. It is mentioned that human learning is a complex phenomenon and learning processes among individual students are never identical (Thurairaj et al., 2010). Hence instructors should vary their teaching methods. In brief, it is possible to implement this idea to improve the English proficiency level of the nations. It is undeniable that Facebook and Twitter could make language learning pleasurable. However, time is needed to find people with good command of the language who are willing to share their knowledge on Facebook and Twitter. Besides, we also need to ensure that everyone who is in sync with this idea will give full support so that social networking sites can remain as dynamic language learning platforms for students.

References

1. Aydin, S. 2012. A review of research on Facebook as an educational. *Educational Communications and Technology* .
2. Drouin, M. 2010. College students' text messaging, use of textese and literacy skills. *Journal of Computer Assisted Learning* , 67-75.
3. Enonbun, O., 2010. Constructivism and web 2.0 in the emerging learning era: A global perspective. *Journal of Strategic Innovation and Sustainability* 6 (4), 17-27.
4. Fodeman, D. &. 2009. The impact of Facebook on our students. *Teacher Librarian* , 36-40.
5. Geraldine Blattner, M. F. 2009. Facebook in the Language Classroom: . *Instructional Technology and Distance Learning (ITDL)* , 17-28.
6. Harkin,J 2009, May 4. Caught in the net. *New Statesman*, 138(4947), 20-26. Retrieved 28 November 2012 from http://www.newstatesman.com/world-affairs/2009/05/online-information-feedback
7. Harrison, R. & Thomas, M., 2009. Identity in online communities: Social networking sites and language learning. *International Journal of Emerging Technologies and Society* 7 (2), 109-124.

8. Kabilan, M. K., Ahmad, N., & Abidin, M. J. Z. 2010. Facebook: An online environment for learning of English in institutions of higher education? *The Internet and Higher Education, 13*(4), 179-187.
9. Kirkwood, K., 2010. The SNAP platform: Social networking for academic purposes. *Campus Wide Information Systems* 27 (3), 118-126.
10. Lim Yung Hui, 2009. Snapshot of social networking in Malaysia. *Grey Review The Social Web Journal.* Retrieved 1 November 2012 from http://www.greyreview.com/2009/07/28/snapshot-of-social-networking-in-malaysia/
11. Madge, C. M., 2000. Facebook, social integration and informal learning at university: 'It is more for socialising and talking to friends about work than for actually doing work'. *Learning, Media & Technology* , 141-155.
12. Melor Md Yunus, H. S. 2012. Integrating Social Networking Tools into ESL Writing Classroom:. *English Language Teaching* , 42-48.
13. Queirolo, J. 2009. Is Facebook as good as face-to-face? *Learning & Leading with Technology* , 8-9.
14. Thurairaj, S. , Mangalam, C. and Marimuthu, M., 2010. Reflection on Multiple Intelligences. *International Journal of African Studies*, 3, 16-25.
15. Thurairaj, S. and Roy,S.S., 2012. Teachers' Emotions in ELT Material Design. *International Journal of Social Science and Humanity*, 2(3), 232-236.
16. Tucker, J. & Courts, B., July 2010. Utilizing the internet to facilitate classroom learning. *Journal of College Teaching and Learning* 7 (7), 37-43.
17. Warschauer, M., & Kern, R. 2000. *Network-based language teaching: Theory and practice.* Cambridge, England: Cambridge University Press.
18. Williams, J. & Chinn, S. J., 2009. Using web 2.0 to support the active learning experience. *Journal of Information Systems Education* 20 (2), 165-174.

Price policy model at the modern shadow market of information technologies

Serghei Ohrimenco, Agop Sarkisian, Grigory Borta

1 Introduction

The topic of underground market of modern information technologies processes modeling, including price policy model, is poorly studied nowadays even though research in this domain is crucial not only to the "official" market, but to the information security experts as well. That is the main reason behind the necessity to discuss the problems of shadow market of information technologies in general, and more specifically the mathematical models behind it.

It is also important to note the fact that this work is a logical follow-up to a number of works performed by the authors and published in the papers.

2 Designation of the main domains of shadow market of information and communication technologies

Shadow economics of certain economical domains (and even countries, regions) traditionally were the subject of economic research. Economics, statistics, law and sociology experts propose a substantial number of definitions which to some degree reflect the state of things in the field unaccounted for at the governmental level. The most common words used to denote this economical activity are "shadow", "criminal", "underground", "intangible", "parallel", "grey", "black", and others. It is important to note the fact that "black" economy is the most autonomous from the legal one, since it produces and consumes the goods that are officially prohibited.

Further we will try to define this economical domain, its structure, objective reasons of its existence, and hypothetical consequences.

We consider that under the term "shadow information economics" one should understand all the individual and collective unlawful activity, related to design, production, distribution, support, and use of components of information and communication technologies that is hidden from society. In other words, shadow information economics is all the criminal information products, services and processes based on IT or using IT. The main economical elements of this domain are unlawful economical relationships, illegal business, which is related to production, distribution and use of prohibited goods and services, sphere of illegal employment. It is important to note the fact that this kind of economics merges unlawful goods and services production, prohibited by national legislations, unlawful sale and purchase of goods and services, and consume of aforementioned unlawful goods and services. Therefore, we can conclude that the main reason of shadow economics existence is a set of conditions that makes it profitable to conduct unlawful activity in the domain of

information technologies. Generalized structure model of shadow information economics is represented on fig.1.

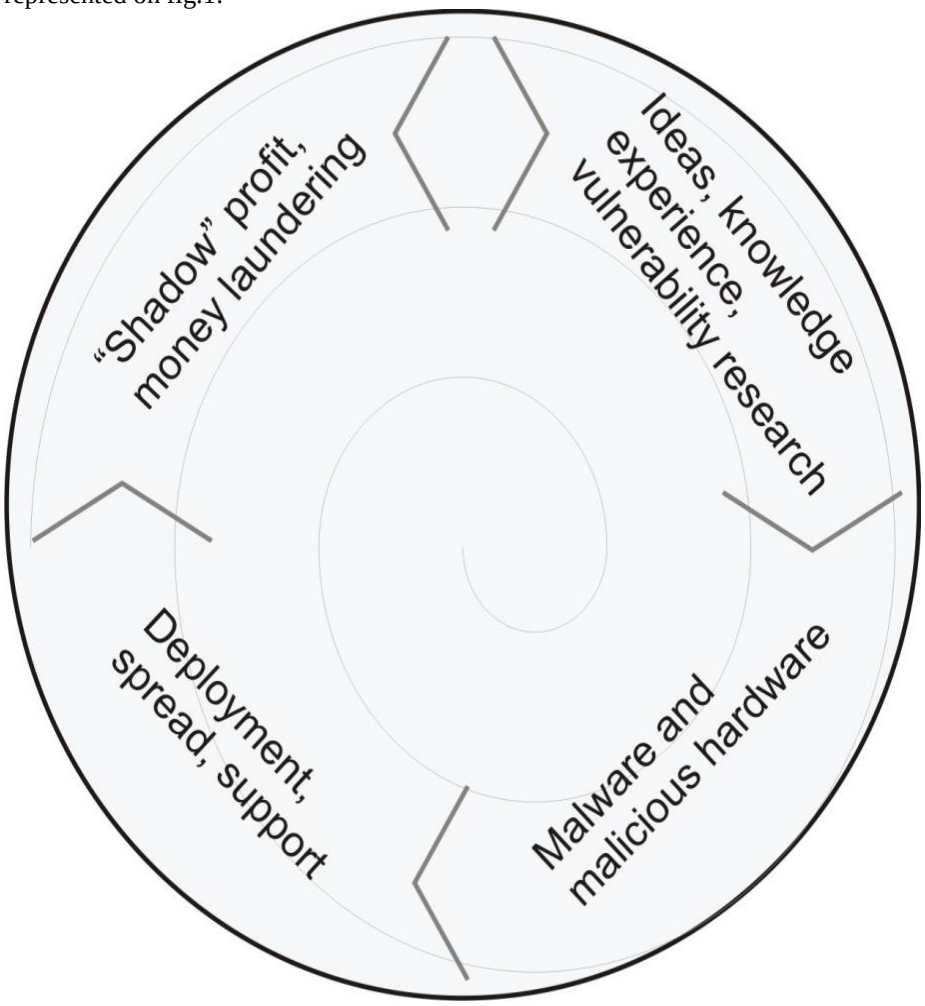


Fig. 1. Generalized structure of the shadow information economics

Let's analyze the subject of shadow information economics in more detail.

First of all, the term is related to a large number of products, that are sold at "black", or "parallel" market of information and communication technologies. The most common among them are the following:

- Development and modification of a large range of specialized software;
- Spyware devices, such as key and electromagnetic impulse loggers, skimmers, wired and wireless sniffing devices.
- Technologies of production and counterfeiting of plastic cards
- Pirating software

Among the most distinct categories of malware we would like to note the following ones:

- Viruses – recently a new kind of viruses started to surface and attract more and more attention. The most common term used to describe them is cyber weapons. The most well-known representatives of this class of malware are Stuxnet, Flame, and DuQu. Their features, ways of spread, malicious activity and payloads differ, but all of them are speculated to have the same origins.
- Spyware – this category may be represented by both software and hardware spies. The cost common examples of the latter are wireless sniffing devices, electromagnetic impulse loggers, etc.
- Trojan horses – due to human factor, this type of malware still remains one of the burning issues of the day.
- Fraudware – also known as rouge security software; is a comparatively recent addition to the malware family. It misleads the user into thinking that it is an anitivirus, while indeed it fakes antivirus activity, such as simulates malware removal, real-time protection, etc.
- Worms – one of the first types of malware in existence is still viable today, though to a smaller extent than before.
- Adware – one of the recent additions to the shadow information market.
- Crimeware – is the type of malware specifically designed to help cybercrime.

One of the most important recent trends is the fact that malware targets mobile devices more and more. Another distinct feature is malware behavior model: it tends to use computer's (or mobile device's) calculating power in order to spread, send spam, conduct DDoS attacks and steals private data rather than tries to cause device malfunction.

The main group of products employed at the shadow information market is represented on fig.2.

Second, this is a large set of criminal services, which is recently often denoted as "Crime as a Service". This kind of service may be used to gain unauthorized access to informational resources, etc.

The most common services of the domain are the following ones:

- Analytics, including search and research of vulnerabilities, market and legislation analysis, etc.
- Identity theft, such as interception and capture of identification information, credit card data, logins, passwords.
- Spam
- Phishing
- Farming
- Extortion
- Sabotage
- Terrorism
- Piracy
- Proxy servers rental
- DoS attacks
- Money laundering, using information technologies
- Botnet creation and rental

The structure of services, provided at the shadow information economics market is represented on fig.3.

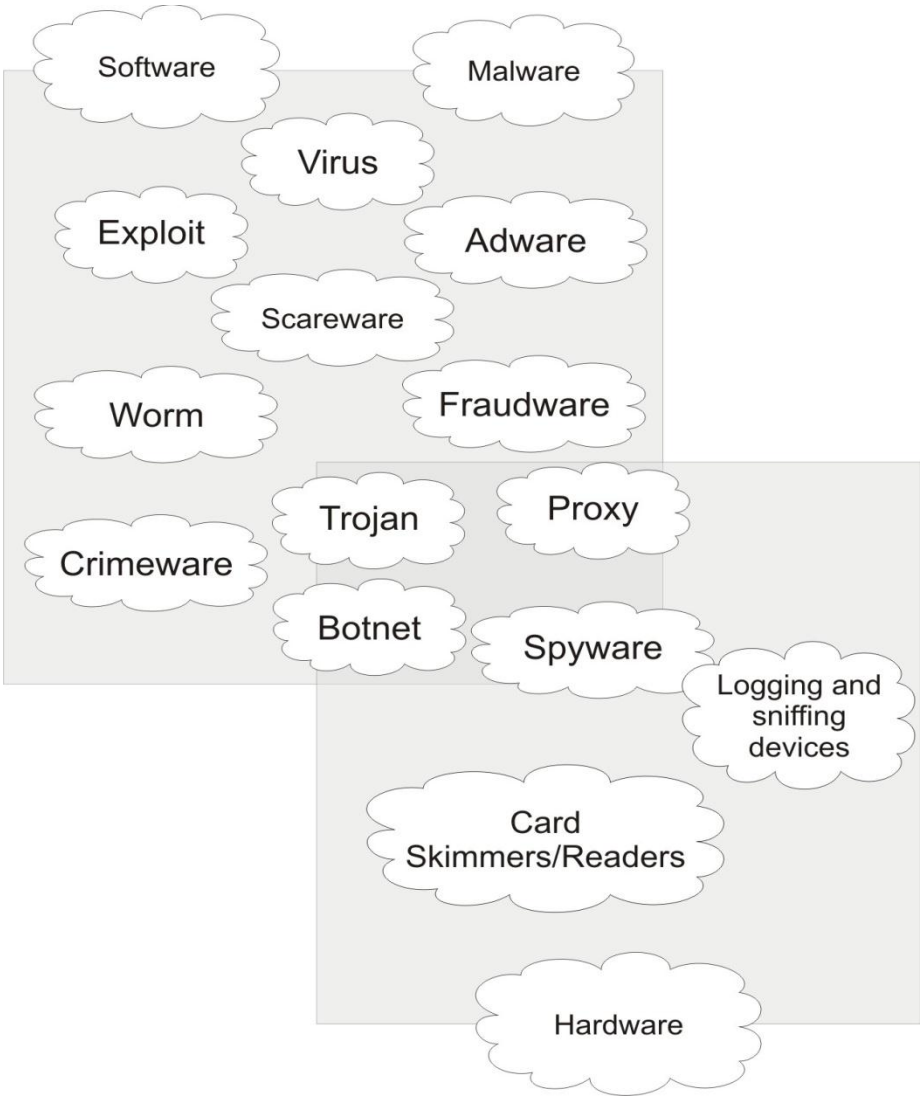


Fig. 2. Products employed at the shadow information market.

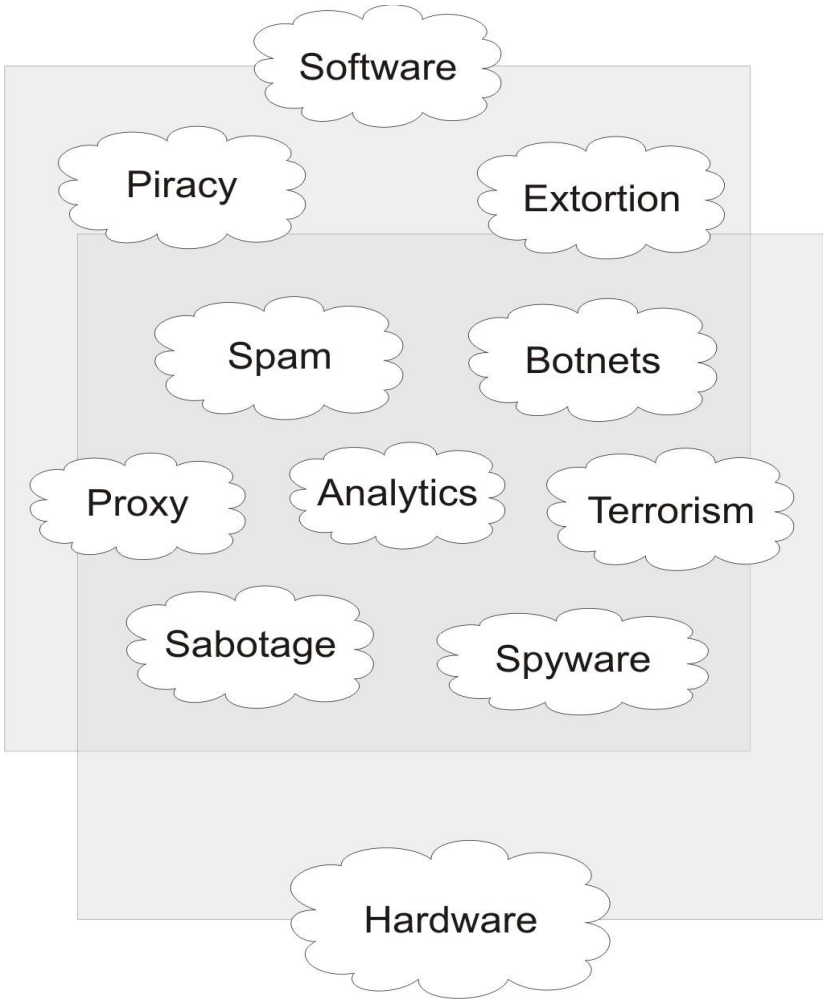


Fig. 3. Services, provided at the shadow information economics market.

3 The main directions of research

According to our opinion, the main strategic subject of research in this domain should be the mechanisms of interaction between the participants of this segment of shadow information market. They may be related to:

- Distribution of goods and services belonging to the shadow information market
- Shaping of demand and supply for goods and services
- Analyzing and choosing a business model
- Establishing a price policy
- Competition

- New participants appearance
- Elimination of participants who failed in the competition struggle, etc.
- This research should provide a solution to the following practical problems:
- Determination and definition of the main distinctive features and properties of shadow products and services.
- Development of mathematical models of interactions between developers, producers, distributors, and users of "shadow" goods and services
- Research of the mechanisms of shadow goods and services market participants interaction.

It is also important to note the fact that the process of price establishment for the "shadow" goods and services has a very important and substantial role for the "shadow" market, because a fast-paced price change and speculative character of use of these changes are noted. Some of the prices are represented in table.1.

Table 2. Vulnerability cost depending on the software.

	Software product	Vulnerability cost
1	Adobe Reader	$5,000 - $30,000
2	MAC OSX	$20,000 - $50,000
3	Android	$30,000 - $60,000
4	Flash or Java Browser Plug-ins	$40,000 - $100,000
5	Microsoft Word	$50,000 - $100,000
6	Windows	$60,000 - $120,000
7	Firefox or Safari	$60,000 - $150,000
8	Chrome or Internet Explorer	$80,000 - $200,000
9	IOS	$100,000 - $250,000

The following factors are to be taken into consideration when a model is developed:

- Aspiration of a seller (distributor, intermediary) to sell their product and for the consumer to buy it
- The market is characterized by presence of active and inactive participants
- Seller might offer a good that he is inexperienced to use and unaware of its realization
- Seller and buyer might not know anything about each other, or know very little.

Price policy model

Let's have a look at a hypothetical economic and mathematic price policy model for an exploit. Preliminary expenses would be as follows:

- Market analysis, including available operation systems, software products, their diversity, popularity, etc.

- Software analysis, including different version popularity, patch frequency and availability, operation system and software architecture, etc.
- Vulnerability analysis, including exploitation possibility and difficulty, etc.

Here are some of the additional conditions:

1. Exploited product popularity – vulnerability prices represented on figure 4 are for the most part due to popularity of the software product under consideration.
2. Zero Day status
3. Exclusiveness – if the vulnerability is being sold to a single buyer, it will surely cost much more.
4. Currency of exploited product version – vulnerabilities that are fixed in newer versions of a product cost less, even if they are not known or published to the public.
5. Exploit stability
6. Application vector – the less actions the victim has to perform, the better
7. Market saturation – obviously, in some countries there are more researchers than in the others. In digital world borders may not be that obvious, but some of the hackers tend to sell their vulnerabilities only to the buyers originating from their own country. That said, in some countries bugs are cheaper than in the others. A good example of a saturated vulnerability market is China
8. Time between launch and response – less actual nowadays than before.

Suppose, that "urgency" of an exploit changes with a constant pace of μ, if at some point in time t a vendor has $Q(t)$ of this product available, then in an interval of time $[t, t+\Delta t]$ the "urgency" will go down to $\mu Q(t) + o(\Delta t)$.

Sales price is constant and equals c. Then, a flow of buyers will be an exponential arrivals with intensity $\lambda(c)$, and in Δt time, $\lambda(c)\Delta t$ buyers will buy an amount of goods equaling $\alpha_1 \lambda(c)\Delta t + o(\Delta t)$.

Hence,

$$Q(t) - Q(t+\Delta t) = \mu Q(t)\Delta t + \alpha_1 \lambda(c)\Delta t + o(\Delta t) \quad (1)$$

Dividing by Δt and limiting transition $\Delta t \to 0$, we get the following differential equation:

$$\frac{dQ(t)}{dt} = \mu Q(t) - \alpha_1 \lambda(c) \quad (2)$$

That needs to be solved with a base condition $Q(0) = Q_0$

The solution to the equation would look like this:

$$Q(t) = \left(Q_0 + \frac{\alpha_1 \lambda(c)}{\mu}\right) e^{-\mu t} - \frac{\alpha_1 \lambda(c)}{\mu} = 0 \quad (3)$$

The salesperson tends to sell his or her exploits (or other malicious software) in the shortest term possible. Let's mark this point in time as T, which gives us the following condition:

$$\left(Q_0 + \frac{\alpha_1 \lambda(c)}{\mu}\right) e^{-\mu T} - \frac{\alpha_1 \lambda(c)}{\mu} = 0 \quad (4)$$

This gives us the sales price:

$$\alpha_1\lambda(c) = \frac{Q_0\mu e^{-\mu T}}{1-e^{-\mu T}} = \frac{\mu Q}{e^{\mu T}-1} \quad (5)$$

The sale price for linear approximation:

$$\lambda(c) = \lambda_0 - \lambda_1\frac{c-c_0}{c_0} \quad (6)$$

This gives us:

$$\lambda_0 + \lambda_1 - \lambda_1\frac{c}{c_0} = \frac{\mu Q_0}{\alpha_1(e^{\mu T}-1)} \quad (7)$$

This, in turn, gives us:

$$c = c_0\left(1 + \frac{\lambda_0}{\lambda_1} - \frac{\mu Q_0}{\lambda_1\alpha_1(e^{\mu T}-1)}\right) \quad (8)$$

Then, sales receipt would be as follows:

$$S = \alpha_1 c\lambda(c)T = c_0\left(1 + \frac{\lambda_0}{\lambda_1} - \frac{\mu Q_0}{\lambda_1\alpha_1(e^{\mu T}-1)}\right)\frac{\mu Q_0 T}{e^{\mu T}-1} \quad (9)$$

If the exploits were purchased via intermediaries to be later sold for their wholesale prices, which we assume to be d, the gain would be estimated as follows:

$$P = S - dQ_0 = c_0\left(1 + \frac{\lambda_0}{\lambda_1} - \frac{\mu Q}{\lambda_1\alpha_1(e^{\mu T}-1)}\right)\frac{\mu Q_0 T}{e^{\mu T}-1} - dQ_0 \quad (10)$$

The maximum amount of income being:

$$P_{max} = \left[c_0\left(1 + \frac{\lambda_0}{\lambda_1}\right)\frac{\mu T}{e^{\mu T}-1} - d\right]^2\frac{\lambda_1\alpha_1(e^{\mu T}-1)^2}{4\mu^2 T c_0} \quad (11)$$

Similar methods may be applied to calculations of botnet functioning efficiency, the latter being one of the most common constituent of shadow information economics.

The struggle against botnets is conducted strategically incorrect, as stated in [1]. Measures taken might shut down a single zombified network or even just its part, while the others remain up and running, and the overall amount of spam doesn't decrease significantly. This can be explained by the fact that all the botnet creators have large amounts of reserve productivity. The reserve can easily be brought to service and the time interval may be as short as the time needed to sign a new contract. This kind of actions may be performed in incredibly short terms, and may be determined only by the economical degree of contract elaboration.

Presence of this amount of surplus productivity greatly reduces the efficiency of end-user means of protection for the workstations, such as antivirus, firewall, etc. Another problem arising is the inefficiency of struggle against botnet control centers. A lot of researches share the opinion on the fact that one of the weakest spots in a botnet is service sales.

4 Conclusion

The main conclusion of this paper is development of a strategy that will propose a solution to the shadow information economics market. The following principles would lie at the basis of the strategy:

- Improvement of legislative base of economic regulation aiming to create conditions that will decrease profitability of unlawful activity to the point where it will not make sense to conduct any.
- Development of a state-level partnership aiming at lowering the level of shadow information economics.
- Creation of a set of new workplaces, tax system reformation, making it much more restrictive in the domain of money laundering, toughening the struggle against corruption.

References

1. Chiesa, Raoul. Cybercrime & underground economy: operating and business model. Flare Network. [Online] July 1, 2010. [Cited: July 22, 2012.] http://www.flarenetwork.org/report/enquiries/article/cybercrime_and_underground_economy_operating_and_business_model.htm.
2. Chiesa, Raoul . Cybercrime: reasons, evolution of the players and an analysis of their modus operandi. Bright. [Online] July 1, 2010. http://flarenetwork.org/report/enquiries/article/cybercrime_reasons_evolution_of_the_players_and_an_analysis_of_their_modus_operandi.htm.
3. Greenberg, Andy. Meet The Hackers Who Sell Spies The Tools To Crack Your PC (And Get Paid Six-Figure Fees). Forbes. [Online] March 21, 2012. [Cited: June 21, 2012.] http://www.forbes.com/sites/andygreenberg/2012/03/21/meet-the-hackers-who-sell-spies-the-tools-to-crack-your-pc-and-get-paid-six-figure-fees/.
4. Greenberg, Andy. Shopping For Zero-Days: A Price List For Hackers' Secret Software Exploits. Forbes. [Online] March 23, 2012. [Cited: July 20, 2012.] http://www.forbes.com/sites/andygreenberg/2012/03/23/shopping-for-zero-days-an-price-list-for-hackers-secret-software-exploits/.
5. Richardson, Robert J. Monitoring Sale Transactions for Illegal Activity. IIMA online . [Online] 2006. http://www.iima.org/CIIMA/13%20CIIMA%206-1%20105-114%20Richardson.pdf.
6. Zorz, Zeljka. How much does a 0-day vulnerability cost? Help Net Security. [Online] March 26, 2012. [Cited: July 20, 2012.] http://www.net-security.org/secworld.php?id=12652.
7. Федотов, Н.Н. Ботнеты процветают. SecurityLab. [Online] August 26, 2012. [Cited: September 10, 2012.] http://www.securitylab.ru/blog/company/infowatch/23945.php.

8. Curry, Sam and Williams, Amrit. The Economics of Cybercrime and the "Law of Malware Probability". RSA. [Online] [Cited: September 10, 2012.] http://www.rsa.com/blog/pdfs/economics_cybercrime.pdf.
9. Бортэ, Григорий. Исследование жизненного цикла вредоносного программного обеспечения. laboratory of Information Security. [Online] [Cited: September 10, 2012.] http://security.ase.md/materials/publications/index.html?artNr=6.
10. Sarkisyan Agop, Ohrimenco Sergey. Training Higher Education Degrees Specialists in the Field of Information Security. Laboratory of Information Security. [Online] [Cited: September 10, 2012.] http://security.ase.md/materials/publications/index.html?artNr=15.
11. Бортэ, Григорий. Теневая Информационная Экономика. Laboratory of Information Security.[Online][Cited:September10,2012.] http://www.security.ase.md/materials/si2011/Security_inform_2011_Pages_94_-_96.pdf.
12. Valentino-Devries, Jennifer, Sonne, Paul and Malas, Nour. U.S. Firm Acknowledges Syria Uses Its Gear to Block Web . The Wall Street Journal. [Online] October 29, 2011. [Cited: July 20, 2012.] http://online.wsj.com/article/SB10001424052970203687504577001911398596328.html.
13. Синцов, Алексей. Стоимость эксплойта... SecurityLab. [Online] June 21, 2012. [Cited: September 10, 2012.] http://www.securitylab.ru/blog/personal/asintsov/23097.php.
14. Охрименко, Сергей and Саркисян, Агоп. Подпольная информационная индустрия. The Bulletin of Transport and Industry Economics. Харків : N 29, 2010, p.19-22.
15. Ohrimenco Serghei, Hârbu Eduard, Bortă Grigorii, Sclifos Constantin, Solonenco Oleg, Levandovsky Vlad, Storoj Oxana, Pavlova Lilia. Information security in SMB. Management and Production Engineering. Bielsko-Biala : 2012.pp 203-220.
16. Boehme, Rainer. Vulnerability Markets. What is the economic value of a zero-day exploit? [Online] http://events.ccc.de/congress/2005/fahrplan/attachments/542-Boehme2005_22C3_VulnerabilityMarkets.pdf
17. Охрименко, Сергей and Саркисян, Агоп. Сенчестият Пазар на Инфомационни Технологии. Приложна информатика и статистика -съвременни подходи и методи. Международна научна конференция. 25-26 септември 2009, Равда, с.131-138.
18. Охрименко, Сергей and Саркисян, Агоп. Экосистема Информационной Безопасности. //Съвременни измерения на търговския бизнес. Комуникация между наука и практика. Юбилейна научнопрактическа конференция с международно участие. 12-13 май 2011 г. , Свищов, с.31-36.

Automation of the system of internal control - reasonableness of the solution

Lilia Pavlova

Academy of Economic Studies of Moldova,
"IT&IS Management" SRL, Republic of Moldova, Analyst of Business Process and Information Technology
wiz_lilia@mail.ru, liliapav1@gmail.com

Abstract. This article addresses the issue of internal control and expediency automation of system of internal control. It is realized the analysis of options for building an automated system of internal control, including the most popular information systems. The paper evaluated the difficulties that may arise in the process of automation, and also highlights factors that should be paid special attention in selecting solutions for automation of the internal control. At the end of are marked benefits of the implementation of information systems of internal control.

Keywords: automation, internal control, risk management, monitoring.

Internal control is always present in activity of companies, often informally and unconsciously. The increasing complexity of management in companies led to a complication of internal control and the fall of its effectiveness using informal or poorly organized approach.

Inspection and testing of controls, formation reporting on internal control is labor-intensive processes, which are time-consuming. Documented and implemented approach to organization of internal control system can be optimized to reduce time spent on repetitive tasks.

The internal control system is most effective when the control is built into the infrastructure of the company and is part of its core business. Execution of actions, provided by the system of internal control, ensures that all transactions are recorded in accordance with the requirements. Development of the system of internal control includes the performance of action for integrating the components of internal control in all business processes related to financial reporting, operation activity, training, asset management and the use of automated systems.

Information technology is a critical element in supporting reliable and effective system of internal control, taking into account the quickly cost recovery on information systems and equipment, as well as the costs and risks related to failure to automate control procedures.

Automation of internal control - is a phenomenon, that gaining more popularity in modern market and providing a business efficiency. It should be noted that automation of internal control does not resolve organizational problems, but is only a tool to reduce labor costs.

Necessity of automation caused by the fact that to satisfy the requirements of COSO (The Committee of Sponsoring Organizations of the Treadway Commission) need to control the risks of all business processes, which is very time-consuming and expensive to do without the implementation and use of integrated information systems for management and control. Automated system should provide assistance in monitoring the changes of the control environment, analysis and evaluation of internal control.

The competent automation system of internal control allows you to effectively manage risks of the company and control them. This process involves the creation of the logical and the information structure for risk management, and only then automation. Note the following options for the construction of an automated system of internal control:

- development and design own system - independent automation, or attraction to this process of third-party developers;
- implementation of existing solutions to automate the systems of internal control;
- building on the existing integrated information system.

Should take into account that every company has historically set of various systems and applications. This factor must be taken into account, not only in creation of a common information space, but also in the automation of internal controls.

In choosing solutions for the automation of internal controls is necessary to generate functional requirements for the system, the package of general and technical requirements, security and monitoring requirements, requirements for integration and interoperability with existing IT infrastructure of the company.

There are many opportunities and solutions for the automation of internal controls. These opportunities can vary depending on the analysis tools, on the implementation of flexible solution and continuous monitoring, implementation of complex integrated systems, and that much important - return on investment.

Among the most popular solutions for automation systems of internal control are:

- Microsoft Office Solution Accelerator for Sarbanes-Oxley (MOSASO), operates on a technological platform of Microsoft;
- modulus of Oracle Internal Controls Manager, part of the Oracle E-Business Suite. Module integrates into a single system components of internal control, which are responsible for documentation, testing, and monitoring of internal controls and compliance with legal requirements;
- the solution of SAP Management of Internal Controls, representing a logical combination of different mechanisms SAP to build an automated system of internal controls;
- the solution of Workplace for Business Controls and Reporting, IBM, allowing company to optimize the control of financial indicators. This solution increases the transparency of the systems of internal control and business processes, allowing companies to more accurately comply with all legal requirements;
- solution Microsoft SharePoint Portal Server, directed at implementation of the law Sarbanes-Oxley, by creating single information space, workflow and tools for monitoring and control;
- solution Sarbanes-Oxley Corporate Assessment Accelerator, developed by Mercury, allows directories of business processes, control objectives, risks, control procedures to reduce the risks of misstatement of the financial reporting and to create different types of reports;

- solution Compliancy's Internal Control Management (ICM) provides automation and risk management in the company, control and analysis of compliance.

Structure of reduction the cost of implementation of the system of internal control and improving the effectiveness of controls with reflection of return on investment is shown in Fig. 1.

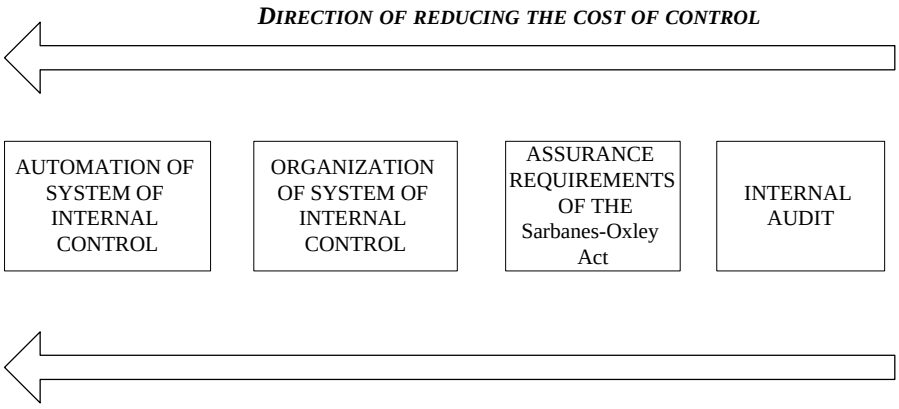


Fig. 1. Direction of increasing the effectiveness of controls.

Necessary to note that the automation of control mechanisms will not only reduce labor costs and the cost of time, quickly analyze and assess the current and potential risks, but will also provide the company's management and reliable information in real time about the state of the system of internal control.

It should be noted that the choice between manual and automated control procedures are preferred automation, due to the fact that this type of control is more effective and efficient, less susceptible to manipulation and reduce the likelihood of errors, independent of human factors.

According to the report about of investigation «Thinking outside the SOX box» of "Ernst & Young", in April 2011, only 3% percent of managers to fully automate most of the key control functions. Over a third of managers noted the use of more than 1,000 controls, and 21% on an ongoing basis using advanced information technologies for the analysis and coordination of activities to ensure compliance with the Sarbanes-Oxley Act.

In process of the automation of systems of internal control may encounter the following difficulties:

- the need to build a conceptual model of internal control in conjunction with the development of technical specifications for the information system;
- consolidation of indicators of the effectiveness of internal controls for the companies having branch network, build related reporting;
- lack of skills, experience and knowledge of the staff in the company in the domain of internal control for participation in the project of automation;

- the need for employees to rethink their activities in terms of risk and implementation.

Automated system of internal control and risk management must include an element of modeling of processes or integrated with the system of business process simulation. This factor will allow to analyze changes in the business processes of the company and to evaluate emerging risks, the need to change the control and monitoring procedures. It is important to ensure ongoing control and monitoring of the execution of control procedures not only automated, but manual.

Automated system of internal control should combine into one system all the components of internal control, such as documentation, monitoring, testing, assessment and reporting, provide the ability to actively monitor the internal control processes of business in real time, and to promptly report about the state of the system of internal control and the need for corrective action to correct deficiencies.

In selecting solutions for the automation system of internal controls must take into account the following factors:

- the ability to integrate with existing in company systems and applications;
- scalability and flexibility;
- efficiency of information exchange;
- simplicity to use;
- visibility of results;
- versatility as applicable to their system of performance, control, planning;
- reducing the cost of controls and eliminating duplication.

Automation of monitoring of internal controls provides efficiency and improved IT management, allowing you to more effectively analyze user behavior in the IT infrastructure, to monitor any changes, to ensure compliance with information security requirements, to control violations and generate the appropriate reports. Automation of processes of testing network equipment for compliance with approved policies also can significantly increase the overall level of information security and the effectiveness of control.

Process of automation of internal controls should provide the following capabilities:

- analysis of the completeness of control and benchmarking in business processes;
- capacity planning and test management of controls;
- optimization of workflow and centralized access to all documents;
- use of effective analytical tools and techniques for early warning for risk management;
- ensuring with more accurate information;
- monitoring of access to data to ensure control over leakage and distortion of data;
- improving the exchange of information about the risks and the critical parameters of the company's activity;
- formation of different statements relating to internal control, using powerful analytical tools and convenient tools;
- ensuring transparency of corporate processes, reporting, risk management and system of internal control.

Application of an automated approach to internal control provides senior management an adequate tool for the analysis and evaluation of factors that may affect the reliability of financial reporting, effectiveness and efficiency of operations, compliance with laws.

Optimization of internal control determines the motion of company from management issues of the actual results to preventive management, providing information in real time and providing continuously monitor the state of control procedures thus preventing possible problems.

Automation of internal control is the key to achieving the highest level of maturity of the system of internal control. This process aims at improving the effectiveness of control procedures and their monitoring in carrying out of business processes.

References

1. Teresa Wingfield: Effective IT Controls: Why Continuous Monitoring Requires Automation. November 2006;
2. Committee of Sponsoring Organizations of the Treadway Commission (COSO): Enterprise Risk Management – Integrated Framework. 2004;
3. Ernst & Young: Thinking outside the SOX box. 2011;
4. Vyacheslav Vetrov: Sarbanes Oxley Act: key moments of creation of the internal control. Edition PC Week/RE (527)17`2006, 16.05.2006;
5. Ernst & Young: Significant opportunities for improvements in the field of compliance with the requirements of the Sarbanes-Oxley Act. 2 September 2011.

Risk based software testing

Oxana Storoj

ASEM, Moldova

Key words: risk, testing, quality, software testing.

Intruduction

Testing any real-world project is a task than cannot be fully done, due to the permanently changing external environment. Management of any system needs to take into account this knowledge and estimate any failure than can happen and would cost a lot in case of its occurrence.

Risk based testing is a testing strategy based on determining potential risks and testing parts of the project that has higher probability of failure and higher financial losses.

It is very important to determine the acceptable level of project quality and to confirm that risks remained after testing are acceptable to the business.

Risk is a probability of appearance of any event or condition that can negatively impact the project.

Risk based testing can be done in two main directions:

- Product quality risks – can make negative output on product itself.
- Planning risks – can have negative influence on overall project.

To start effective risk based testing it is important to estimate every risk and to start testing that parts that have higher risk severity and priority.

Risk-Based testing process

Any testing process is staring with identifying main goals and testing aids. In case of risk based testing we can say that main goal is improving product quality by determining all risks and reducing probability of their appearance. Testing process can consist of next steps.

- Making list of risks and identifying their priority.
- Producing tests oriented to research of each risk.
- In case of new risk appearance testing scope is updating and new tests are producing.

This process will stop in case if required quality level is achieved (Fig.1. “risk-based testing process”).

All risks should be linked to specific product specifications. If any specification doesn’t have any risk it is necessary to find out why. It can be that functionality is of low importance. Or functionality can be little-known and it is impossible to identify any risks. In second case QA engineer has to examine mentioned functionality to find risks.

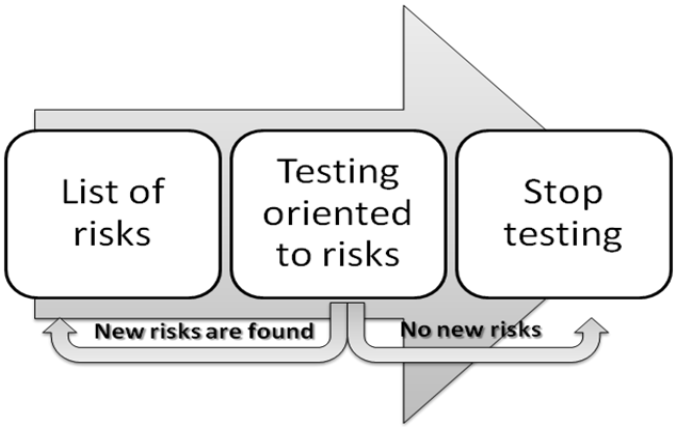


Fig.1. "Risk-based testing process"

First step "Making list of risks" is more important than the other steps. The process of risk identifying and analyzing consist of some more steps.

1.Identifying the quality risk analysis team.
2.Selecting an analising method (informal analyzing, according to ISO 9126, cost of failure, detailed risk effect etc.).
3.Identifying the quality risks: prioritizing, selecting appropriate risk mitigation actions. It may include also reviews of requirements, design, code, programming techniques and other documents. If such problems were identified they should be immediately resolved.
4.Review, revise, and finalize the quality risk analysis document.

Thereby testing scope and man-hours will change from one release to other. It is important to mention that in reality not only absence of new important risks can stop testing process. There can be many other reasons, for example:

- The most important bugs and first-priority risks were fixed;
- Delivery date is too close;
- Budget was used up;
- Project is no more in use, etc.

Risks identification

Not all risks are equal and there are a number of ways to assess the level of risk. So, it's very important to know how to identify risks, to assess their level, to implement appropriate tests and other controls, and to report test results based on risk. Risk identification involves collecting information about the project and classifying it to determine the amount of potential risk in the test phase and in production.

The first thing we should do during risk-based testing is to priorities all risks.

Priority (urgency of the risk). The level of priority indicates whether the risk must be tested or can be postponed to next releases. The priority status is set based on the customer requirements. Priority might change according to schedule, requirements and test plan.

Priority should be set and should reflect importance for testing. According to experience of many testing teams it is enough to define 3 or 4 priority levels. It depends on project size. I will propose four levels.

1.Critical risk – in case of its appearance it blocks some parts of testing without workaround. It is very visible to customer.

2.High risk – provoke appearance of very bad defect but with workaround. Many customers would complain about the issue. The issue does not conform to what was stated as a requirement for the release.

3.Medium risk - all other risks when workaround exist or when impact on testing is very small. Some customers are impacted by it but there is a workaround. Issue than can appear affects not critical areas of the system

4.Low risk – in case when defects that can appear are cosmetic or trivial. Only few customers even notice it much less are impacted by it (i.e., not very visible or detrimental)

Risk traceability matrix can be used for risk estimating. Intersection of probability of occurrence and impact of occurrence gives us the risk severity and priority.

$$Risk = Cost\ of\ Failure = Probability\ of\ Occurrence \times Impact\ of\ Occurrence$$

Below traceability matrix is presented.

Probability \ Impact	Low	Medium	High	Critical
Critical	High	High	Critical	Critical
High	Medium	High	High	Critical
Medium	Low	Medium	High	Critical
Low	Low	Low	Medium	High

Matrix 1. “Matrix of risk priority”

After all risks that were found have been prioritized QA team can start testing. As a rule it starts with those areas that have critical risks. Bugs that are found are fixed from critical to low.

Risk analyses methods.

There are a plenty of different methods of prioritizing risks. Selecting one of them depends on project type, time for testing, team experience, project budget and other factors. Typically, the risks are grouped or organized by major risk categories, such as functionality, performance, security, and so forth. Once the risks are identified, each risk is assigned a level—a measure of its degree of importance and priority, as was mentioned above. Following are some methods for analyzing risk.

1. Brainstorming (or informal). It is a simple but effective attempt to help people think creatively in a group setting without feeling inhibited or being criticized by others. Such method provide an easy way to get started in quality risk analysis. Is used mostly in formative project planning and can also be used to identify risk scenarios for a particular project. The intent is to encourage as many ideas as possible, which may in turn, trigger the ideas of others.

2. According ISO 9126 quality risks analysis methods should describe characteristics and sub characteristics of system quality standard. The six main characteristics of quality are: Functionality, Reliability, Usability, Efficiency, Maintainability, and Portability. In each category there are two or more sub characteristics.

3. Cost of failure quality risk analysis method focus on the following question: What are the expected losses associated with various risks, and how much should one spend to reduce those risks? An expected loss is the product of the probability of the loss multiplied by the cost of the loss. Such techniques allow the project management team to make a hard-nosed, economic decision about testing.

4. Probability analysis method. QA team should specify a probability distribution for each variable, and then consider situations where any or all of these variables can be changed at the same time. Defining the probability of occurrence of any specific variable may be quite difficult, particularly as political or commercial environments can change quite rapidly.

5. Delphi method. The basic concept is to derive a consensus using a panel of experts to arrive at a convergent solution to a specific problem. This is particularly useful in arriving at probability assessments relating to future events where the risk impacts are large and critical. The first and vital step is to select a panel of individuals who have experience in the area at issue. For best results, the panel members should not know each other identity and the process should be conducted with each at separate locations. The responses, together with opinions and justifications, are evaluated and statistical feedback is furnished to each panel member in the next iteration. The process is continued until group responses converge to s specific solution.

Conclusions

Risk based testing helps to make system better by finding the most critical areas and the most critical failures and fixing them. It is a powerful testing technique that enables the QA teams to streamline their testing efforts. Organizations that implement this technique are in a better knowhow of the risks that are inherent in their applications and of the risks that are actually significant. It allows QA teams to make informed decisions while setting a clear test exit criteria.

Risk Based Testing is an approach that requires skill and experience to isolate the most important tests on the basis of technical and business constraints.

Literature:

1. http://coach4qa.com/тестирование-на-основе-рисков/
2. Article by Rex Black "Quality Risk Analysis", 2003.
3. Presentation by CIBC – TASSQ - "Risk Based Testing. Why We Need It & How To Do It", November 2010.
4. Boris Beizer, "Software Testing Techniques" Second Edition, 1990.

Primary Study on the Relationship of Income Distribution and Economic Growth

Shi Yizhe[1]

[1] Sofia University "St. Kliment Ohridski"
Sofia, Bulgaria
Shi.Yizhe@abv.bg

Abstract. What kind of influence will income distribution have on economic growth? This has always been an important question in economics. Theoretical analysis will be made on transmission mechanism of income distribution—consumption demand—economic growth indicating that income distribution functions on economic growth through consumption demand. Empirical analysis is conducted on the relation of consumption demand and economic growth, income distribution and consumption demand with related economic data of China from 1990 to 2007. Result of analysis shows that income distribution adjustment will be significant to long-term stable growth of economy. On this basis, author proposes the classification and selection scope of income distribution policy, and simulates the effect of some income policy tools.

Keywords: Income distribution, economic growth, consumption demand.

1 Introduction

After reform and opening, China economy has been developing rapidly and steadily for over 30 years. However, meanwhile, conditions of income distribution are going worse. Especially after 1990s, Gini coefficient often exceeds international warning limit. Therefore, what kind of influences will the sharply increasing income gap have on economic development of China in the future? This has become a question to be answered in theory and reality.

Basic on the theoretical analysis of transmission mechanism of income distribution-consumption demand-economic growth, author has made an empirical study on their relation by the data from 1990 to 2007. According to the conclusion, author proposes suggestions hoping to increase consumption demand and promote economic growth.

2 Building Basic Model

Firstly, we need to establish a basic model to study the influence of income distribution on economic growth through consumption demand.

Supposing, the society is divided into n levels according to income. Per capital income of each level is expressed degressively as R_i (i=1,2， 3， ……n); average propensity to consume of this level is c_i, the number of people in this level N_i, consumption demand of

whole society C. Then：

(1) $C = \sum_{i=1}^{n} R_i c_i N_i$

N is total population, so

(2) $N = \sum_{i=1}^{n} N_i$

Y indicates national income, I investment, G government purchase, X-M net export, Y/N per capita national income. Economic growth is measured by increment of per capita national income.

(3) $$\frac{Y}{N} = \frac{C+I+G+(X-M)}{N} = \frac{C}{N} + \frac{I}{N} + \frac{G}{N} + \frac{X-M}{N}$$

Substitute formula (1) and (2) to (3) and t_i indicates the proportion of level i in total population.

(4) $$\frac{Y}{N} = \sum_{i=1}^{n} R_i c_i t_i + \frac{I}{N} + \frac{G}{N} + \frac{X-M}{N}$$

It can be seen from formula (4) that population proportion t_i and income level R_i of each level will affect the level of economic growth. Both are just important indicators to measure the status of income distribution.

Through the analysis of above theoretical model, it may be seen that income distribution status will have influence on economic growth through the transmission mechanism of income distribution-consumption demand-economic growth. In the following, author will make empirical analysis of this transmission mechanism, and simulate the influence of income distribution on economic growth by adjusting the policies of income distribution.

2 Empirical Analyses of Consumption Demand on Economic Growth

In order to analyze the influence of income distribution on economic growth, let's take a look at the influence of various demands, especially consumption, on economic growth from 1990 to 2007. This paper introduces contribution rate to express the influence of various demands on economic growth. The formula is as follow:

(5) $$\nabla_C = \frac{C_n - C_{n-1}}{Y_n - Y_{n-1}} \qquad \nabla_I = \frac{I_n - I_{n-1}}{Y_n - Y_{n-1}}$$

$$\nabla_G = \frac{G_n - G_{n-1}}{Y_n - Y_{n-1}} \qquad \nabla_{X-M} = \frac{(X-M)_n - (X-M)_{n-1}}{Y_n - Y_{n-1}}$$

Thereinto, ∇_C ∇_I ∇_G ∇_{X-M} separately indicates the contribution rate of resident consumption demand, investment, government purchase and net export to economic growth;

Y indicates GDP figured out by expenditure method; subscript n expresses economic variable of current year; subscript n-1 expresses the economic variable of last year.

According to formula (5), we have:

Table 3. Contribution rate (%) of resident consumption, investment, government purchase and net export to economic growth from 1990 to 2007

Year	Resident Consumption	Investment	Government Purchase	Net Export
1990	37.36	1.8	10.44	50.30
1991	49.57	24.30	15.53	10.50
1992	54.79	34.20	17.71	-6.70
1993	44.59	78.60	14.91	-38.10
1994	22.56	43.80	7.64	25.90
1995	34.51	55.00	10.19	0.30
1996	46.47	34.30	13.63	5.60
1997	28.38	-7.40	8.62	70.40
1998	43.42	29.30	13.68	13.60
1999	57.87	52.80	18.93	-29.60
2000	47.56	21.70	16.24	14.40
2001	36.79	50.10	13.21	-0.10
2002	31.97	48.80	11.63	7.60
2003	25.90	63.70	9.40	1.00
2004	28.38	55.30	10.32	6.10
2005	26.41	38.10	9.69	25.80
2006	28.09	41.74	10.66	19.51
2007	30.87	40.91	12.04	16.17
Total	34.36	42.73	13.33	9.58

Data source: calculated by author basic on data of "China Statistical Year Book 2008"

It can be seen from table 1 that it matches with our theoretical analysis of various demands. In the economic development of China for nearly 20 years, investment demand makes the greatest contribution; the second place is resident consumption demand; purchase demand of government makes outstanding contribution only when government expands the scale (before and after 1991) and implements expanding financial policy (1998-2000). Contribution of export fluctuates greatly and is lack of principle. But, after 2000, the contribution rate of resident consumption demand drops rapidly while investment demand rises sharply, which indicates that the rapid economic growth of China results in great amount of investment. On the other hand, it also indicates domestic consumption demand increases slowly and its influence on economy is gradually being weakened. As mentioned by the above theoretical analysis, due to the instability of investment demand and stability of consumption demand, Chinese economy cannot just rely on the drive of investment demand to embody sustainable development, but should fully exert the function of consumption

demand on economy. As said before, income distribution condition will make important influence on consumption demand. Next, this paper will further study the influencing mechanism of income distribution for consumption demand and its effect.

3 Empirical Analyses of Influence of Income Distribution on Consumption Demand

Income distribution condition of a society can be expressed by the income level of each class classified by income and the population proportion of each class in total population. Formula (4) indicates that the alteration of population proportion of each class and income level will affect consumption demand and then the growth level of economy.

We adopt the percentage of alteration of Gini coefficient as the indicator to measure the change of income distribution condition; percentage of alteration of social average consumption trend as the indicator to measure the change of consumption demand. Then:

$$(1)\ \mu_i^C = \frac{C_i - C_{i-1}}{C_{i-1}} \qquad \mu_i^{GINI} = \frac{GINI_i - GINI_{i-1}}{GINI_{i-1}}$$

Thereinto, μ_i^C indicates the proportion of alteration of average consumption trend in phase i; C_i indicates the average consumption trend of phase i; C_{i-1} the average consumption trend of phase i-1; μ_i^{GINI} the proportion of alteration of Gini coefficient in phase i; $GINI_i$ the Gini coefficient of phase i; $GINI_{i-1}$ the Gini coefficient in phase i-1

As China statistical department made statistics of average income and consumption by urban and rural population, there is no available data of average consumption trend of each phase which can be directly applied to calculation of model. For that, author calculated the average consumption trend of whole society (see table 2). The calculation formula is as follow:

$$(2)\ C = \frac{C_u P_u + C_r P_r}{R_u R_u + R_r P_r}$$

Thereinto, C_u is per capita nonproductive expenditure of urban population; C_r per capita life consumption expenditure of rural residents; R_u per capita controllable income of urban residents; R_r per capita net income of rural residents; P_u urban population; P_r rural population.

Now, author will establish another linear model:

$$(3)\ \mu^C = \beta_0 + \beta_1 \mu^{GINI} + \varepsilon$$

Thereinto, μ^C proportion of alteration of average consumption trend; μ^{GINI} proportion of alteration of Gini coefficient; ε error term.

Table 2. Average consumption trend (%) and Gini coefficient of China from 1990 to 2007

Year	Average Consumption Trend	Gini Coefficient	Year	Average Consumption Trend	Gini Coefficient
1990	95.75	0.3484	2000	77.47	0.4170
1993	82.69	0.4196	2001	76.01	0.4310
1994	82.36	0.4334	2002	76.88	0.4440
1995	82.80	0.4151	2003	75.99	0.4580
1996	81.29	0.3980	2004	75.68	0.4650
1997	79.38	0.3979	2005	76.51	0.4700
1998	77.04	0.4030	2006	75.34	0.4960
1999	75.74	0.4164	2007	73.96	0.4800

Data source: calculation of author basic on "China Statistical Yearbook 2008"

Basic on data of Table 2, make regression analysis on model (8) with SPSS statistical software. Its result is shown in table 3.

Table 3. Regression result of influence of Gini coefficient on average consumption trend

Explanatory Variable	Variable Value	T Value	R^2	F Value
Constant Item	-4.88E-03	-0.769	0.639	22.818
Gini Coefficient	-0.500	-4.777		

Regression result of model (8) shows that from 1990 to 2007, income distribution condition of China can explain the change of consumption demand to a great extent. Every 1% of increase of Gini coefficient, average consumption will decline 0.5%. It can be seen that aggravation of inequality of income distribution will result in the decline of social average consumption, which fits the conclusion of our theoretical analysis.

Through empirical analysis of income distribution and consumption demand, consumption demand and economic growth, it can be seen that income distribution, consumption demand and economic growth forms a close relation. By carrying out specific income distribution policy, consumption demand can be changed affecting economic growth. The next section of paper will classify various income distribution policies and simulate the implementation effect of policy.

4 Classification, Selection and Simulation of Policies to Adjust Income Distribution and Promote Economic Growth

It can be seen from formula (4) that there are 3 main factors which can affect consumption

demand and economic growth. Per capita income of each class R_i, average consumption trend of class c_i, population proportion of this class t_i. Per capital consumption demand will affect per apital GDP. Formula of per capital consumption demand is as follow:

$$(1)\ \frac{C}{N} = \sum_{i=1}^{n} R_i\, c_i t_i \quad (9)$$

According to formula (1), the hypothesis of per capital income of each class ranking from high to low, for any $1 \le i \le j \le n$, there is $R_i > R_j$ (10);

According to the hypothesis that marginal consumption tends to decrease progressively, it can be known that average consumption of high-income class tends to be lower, so $c_i < c_j$ （11）

And, $R_i\, c_i > R_j\, c_j$ （12）

Because $\sum_{i=1}^{n} t_i = 1$, supposing that the proportion of population of different income classes in total population is changing meanwhile, any decline of t_j is necessarily relative to the improvement of t_i. While other conditions are unchanged, combining formula (12), it can prove that the change of this kind of income distribution will increase per capita consumption and national income. It means improving the income of low-income population and lowering down the proportion of low-income population will promote economy. Due to the hypothesis that per capita income of other classes is unchanged, this method of lowering down the proportion of low-income population cannot be embodied by transfer payment policy, but by expanding the employment channel and reducing employment barriers for low-income class. Of course, while the income of other classes is unchanged, the increment of income of any class will promote economy. This way to promote economic growth by increasing the income of each class, especially low-income class, is called income growth policy.

If the proportion of each social class is unchanged, as well as total social income, supposing that income distribution proportion changes at the same time in 2 classes, any improvement of R_j will necessarily be relative to the decrease of R_i. While other conditions are changed, it can be proved that this kind of income distribution alteration will cause the increase of per capita consumption and resident income. Any action of transferring income from higher classes to lower classes will promote consumption and economy. Different from income growth policy, this income distribution policy may affect the enthusiasm of high-income class to provide production factors as it lowers down their income; meanwhile, as the income of low-income class is no longer related with the factors they provide, it may also reduce the enthusiasm of low-income class to provide production factors, and finally results in the decrease of total social income and has negative function on economy. This economic promotion way by transferring the income of high-income classes to low-income classes is called as income transfer policy.

Supposing that the income and population proportion of each social class is unchanged, the increased consumption trend of each class may add to the per capital consumption of society and promote economy. If c_i and c_j receives same-proportion improvement, we have no way to directly determine which class has greater influence on

economic growth when its consumption trend is improved. This lies in the product of per capital income and population proportion of this class. It means that if the population of some class is large enough, improvement of consumption trend of this class will have great influence on economy. This way of promoting economic growth by changing the consumption trend of each class, especially the class with larger population, is called as consumption promotion policy.

Income growth policy and transfer policy form two types of distribution policy. The implementation of consumption promotion policy needs to rely on some income distribution policies to be embodied, so we will subject it to the study of income distribution policy.

Tools of income growth policy include: expand the employment channel of low-income class, provide low-cost and free training and education for low-income class, reduce employment barrier, and promote the flow of low-income group in areas, industries, cities and countries.

Tools of income transfer policy include: progressive personal income tax system, transfer payment system from high-come class to low-income class, property tax system including inheritance tax and granting tax, social security system of a larger coverage, and so on.

The tools of policy which is related with income distribution policy in consumption policy include: perfect social security system (lowering down the preventive deposit of each income class, especially medium-low-income class to improve the trend of marginal consumption), clear and effective private income and property protection system (this system often matches with personal tax declaration system and can raise the consumption trend of high-income class when they are no longer afraid of exposing the fortune.

After defining the classification of income distribution policies of promoting economic growth, we will choose several policy tools to simulate the effect hoping to show that correct income distribution policy will have influence on economic growth. All the policy simulation is basic on related data of "China Statistical Yearbook 2008". On the premise that consumption multiplier and income consumption flexibility is excluded out of consideration, we figure out the increment of per capita GDP due to the change of consumption.

In the policy of income growth, we choose the policy of promoting the flow of rural labor to simulate the effect of policy. If this policy tool can successfully transfer the labor of 1/10 rural low-income class (about 20% of rural population) to towns and cities to become the labor of urban low-income class, it is equal to transferring 2% of rural labor forces and the per capita GDP of the whole society in that year will increase by 41.66 yuan.

In income transfer policy, we choose the policy tool of transferring the income from high-income class to low-income class to simulate the effect of policy. If this policy tool can successfully transfer 1/10 of income of top income class (about 10% of township population) of cities to bottom income class (about 10% of township population), it is equal to that 20% of urban population embody payment transfer; the per capita GDP of whole society of towns, cities and countryside in that year will increase by 27.05 yuan.

In consumption promotion policy, we choose the policy tool of perfecting society security system to simulate the effect of policy. If this policy can increase the average consumption tendency of urban (about 10% of township population) and rural (about 10% of township population) medium-low-income class by 10%, it is equal to lower down the level

of preventive deposits of 20% of total population and the per capita GDP of the whole society in that year will increase by 89.99 yuan.

CONCLUSION

Income distribution will affect consumption demand and economic growth. Facts of economic growth in China indicate that the promotion function of consumption demand to economic growth is weakened, which may be disadvantageous for the long-term stable growth of Chinese economy. Increasing imbalance of income distribution is the important reason for decline of consumption demand. Therefore, government is necessary to correctly distinguish and choose proper income distribution policy tool to embody the sustainable and long-term economic effect.

References

1. Keynes Employment .interest and Money[M] Beijing Commercial Press.
2. Liu Zhenbiao The empiricial analysis of influence to income distribution on economic growth[J] Chen Zhen University 2005-9.
3. Yang Tianyu income distribution and effective demand[M] Beijing Economic science Press 2003.
4. Kiminri Matsuyama The rise of mass consumption societies [J] Journal of Political Economics 2002 (5).

The Analysis of Influence of FDI on Income Distribution Based on Specific Factor Model

Shi Yizhe[1]

[1] Sofia University "St. Kliment Ohridski"
Sofia, Bulgaria
Shi.Yizhe@abv.bg

Abstract. Author of this paper introduced a specific factor model to analyze the short-term influence of foreign and domestic capital on income distribution as a special factor. Theoretical analysis and result of empirical inspection show that foreign capital which contains technical advantages is one of the factors for enlarged income distribution difference in China. And then author made further explanation of inspection result and drew the conclusion and enlightenment.

Keywords: FDI, income distribution, specific factor.

1 Introduction

The enlarged income distribution gap of China has attracted the attention of the world in recent years. According to official statistics, Gini coefficient of China increases to 0.40 in 1994 from the 0.33 of 1980 and then 0.46 in recent years. Current income distribution gap has exceeded India and Ethiopia and the matter of income distribution gap has gradually become an important restriction of China's economic development and social stability. For that, domestic and overseas scholars have made analysis and research from different angles.

Author has analyzed the influence of foreign direct investment (FDI) on income distribution basic on the model of specific factors. The short-term relation between foreign capital as a specific factor and unfair income has been validated showing that FDI really has influence on unfair income distribution to a certain extent. The internal cause lies in the technical advantage of foreign capital. With enlarged technical spillover effect and human capital, income distribution should tend to shrink. When a new technology appears, the capital containing it is in possession of the monopolistic advantages resulting in further expansion of income gap. Seeing from long-term, the influence of foreign capital on income distribution fluctuates.

2 Specific Factor Model

The so called specific factor refers to the kind of factor which is only applicable to one department rather than others within a short term. The short-term and long-term mentioned here is not defined by time, but the mobility of production factors. Under long-term

conditions and supposing production factors are homogeneous, the production factors can be shared by different departments due to homogeneity. That means under long-term conditions, production factors may flow freely between departments. Model of specific factor is mainly applied to explain the influence of changed price of commodity produced by one department on the income distribution of two departments within a short term.

Hypothesis of specific factor model: a country produces two products (x, y) with capital and labor; returns of scale are constant; labor is homogeneous and can flow freely between two departments; total labor amount is fixed and workers are sufficiently employed; capital is a specific factor, which cannot be shared by two departments; investment of each department is constant; all commodity markets and factor markets are entirely in competition.

As we know, price of production factors is determined by the product of marginal value of each production factor. While market price is constant, the price of two production factors is determined by their own marginal material products. As capital is constant, marginal labor productive power of two departments only rests with labor input. Given the premise that commodity price is constant, labor reward is in reverse proportion with labor input. When two departments have the same labor reward, the distribution of labor between 2 departments reaches to equilibrium. Due to various reasons of trade, if P_x (the price of x) rises, when labor reward of X department exceeds Y department, labor is transferred to x department from Y department. According to the rule of marginal return, when P_y rises, it will finally reach to a new balance. The result is that after trade, the income level of this country will rise, but the extent of increase will be lower than the extent of increased commodity price. The return for owner of specific factor is a different case. With the flow-in of labor, x department has increased marginal productivity of capital, so actual income of X department is increased. The case is opposite for Y department. Due to flow-out of labor, marginal productivity of capital decreases and income is lowered down. This is how specific factor model functions on income distribution.

3 Verification of Theory of Income Distribution and Empirical Facts

According to theory of FDI, multinational corporations usually have some kind of monopolistic advantages and are superior to some enterprises in developing countries in technical level and operation efficiency. The technical advantages of foreign-funded enterprises are mainly represented in advanced machines and management. Host countries of investment make regulated regulations for the entry of foreign capital, which restricts foreign capital from accessing to all fields. Meanwhile, due to "internal" production of foreign capital, the flow of domestic capital is also limited.

Usually, when FDI enters the export industry of one country, it will promote the export of this industry and lower down the export of other industries. Rybczynski theorem also indicates that the entry of direct investment will assemble the factors of other industries in current industry and increase the output and export of industry. This principle has been proved in China. According to statistics, the proportion of export/import amount of foreign-funded enterprises in total export/import amount of the whole country increases to 57.43% in 2004 from the 4.04% in 1986. There is also the trend of further expansion.

Due the unique characteristic of FDI and lower labor cost in China, products are highly competitive in overseas and domestic markets. Compared with domestic capital, they can

obtain a higher price. Even though the price is the same, foreign-funded enterprises can still have higher profit due to lower cost. As most foreign capitals come from developed countries, they will continue to adopt the salary system of parent companies. The labor income of foreign-funded departments will be higher than that of domestic capital. According to model of specific factor, labor will transfer from foreign-funded departments until the labor income of both departments is equal. At this time, capital return of domestic-funded departments will be lowered down.

The basic model with foreign capital and domestic capital as specific factors is as follow:

Premise: 1) one country produces 2 products with 2 production factors (the product of domestic department D and foreign-funded department F); 2) scale returns are constant; 3) labor is homogeneous and can freely flow between 2 departments; 4) total labor amount is constant and workers are sufficiently employed, namely $\overline{L}=L_x+L_y$ (unemployed population can be regarded as newly-added population); 5) capital is a specific factor and cannot be shared by two departmetns due to limitation of technology and policy; 6) investment of each department is constant; 7) all commodity markets and factor markets are fully open to competition.

According to premise, the production functions of two products can be expressed as follow:

$$F=f(K_f, L_f)$$

$$D=f(\overline{K_d}, L_d)$$

As shown in fig 1, under the condition of complete competition, if capital efficiency is equal, balance point lies in E and the factor return of two departments is separately:

$$w_f=P_f \cdot MP_{Lf}, r_f=P_f \cdot MP_{kf}$$

$$w_d=P_d \cdot MP_{Ld}, r_d=P_d \cdot MP_{kd}$$

1) Empirical inspection of specific factor return. As the concrete data of profit of foreign-funded department is hard to be accessed, we analyzed the profit-gaining status of foreign-funded department by the proportion of production value of foreign-funded enterprises in total production value and tax of foreign-funded enterprises in total revenue.

Data of Table 1 reflects the tax preference for foreign-funded enterprises; on the other aspect, it can be seen that the proportion of production value created by foreign capital is increasing. Meanwhile, tax also follows the same trend. We can say that the income of foreign-funded department is increasing. On the contrary, the proportion of domestic-funded department is decreasing. This fits the analysis of specific factor model on income of specific factors.

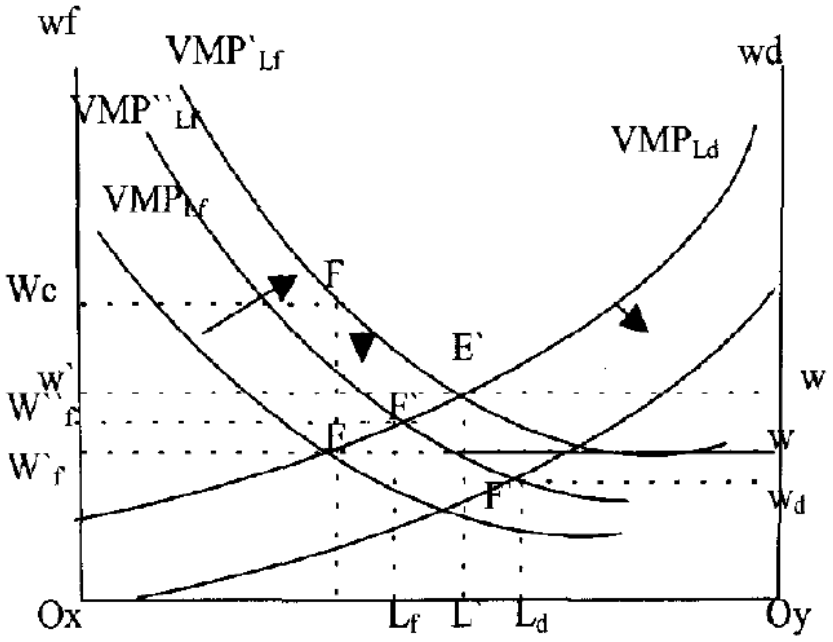


Table 4. Proportion of production value of foreign-funded enterprises and revenue from 1992 to 2002.

Year	Foreign-funded production value/total production value (%)	Tax of foreign-funded enterprises/total revenue (%)
1992	7．09	4．25
1993	9．15	5．71
1994	11．26	8．51
1995	14．31	10.96
1996	15．14	11.87
1997	18．57	13．16
1998	24．00	14.38
1999	27．75	15．99
2000	22．51	17．50
2001	28．05	19. 02
2002	33．37	20. 52

Data source: website of statistical department of PRC (www. stats. gov. cn)

2) Empirical inspection of labor salary

Table 2 average salary of industries in some areas from 2001 to 2002 (unit: yuan)

Area	Year	state-owned units	township collective units	Foreign-funded units
Nation	200l	11178	6867	1810l
	2002	12859	7667	7892
Beijing	2001	19776	11063	34481
	2002	23754	11997	39428
Tianjin	200l	15110	8345	15564
	2002	17059	9350	17643
Shanghai	2001	21961	13693	28787
	2002	24719	14851	30192
Guangdong	2001	16779	9040	18192
	2002	19696	9881	19323

Data source: website of statistical department of PRC (www. stats. gov. cn)

From the nation, average salary of employees working for foreign-funded enterprises is obviously higher than state-owned and collective enterprises. Developed areas also report the same result. According to model of specific factors, this kind of unfairness only happens in owners of specific factors. With the free flow of labor from domestic-funded department to foreign-funded department, the unfairness of labor income will disappear. This does not conform with empirical inspection.

3) Explanation of empirical inspection. Model of specific factors includes the basic hypotheses that labor is homogeneous. But the fact is not like that. Labor is non-homogeneous and can be divided into human capital and non-human capital, skilled and non-skilled labor. As foreign capital is in possession of technical advantages, foreign capital is a kind of effective capital and needs to combine with human capital and requires skilled labor. In the early stage of reform and opening, foreign companies mainly invested in labor-intensive industries and the demand for skilled labor was not significant; since 1990s, foreign companies gradually increased the investment in capital-intensive and technology-intensive

industries and raised the demand for senior technical and management talents, which brought great impact to domestic enterprises. Under this impact, domestic enterprises had to reduce the number of employees to increase the efficiency in order to survive and develop in the market. Consequently, large-scale unemployment tide broke out in the middle of 1990s worsening the unfairness and poverty in the cities. Foreign-funded department also attracts the labor factors of other departments to increase the labor supply. In this situation, salary level tends to decline. However, the salary level of departments where labor flows out does not necessarily rise because the loss of core talents lowers down the efficiency of production and operation, and may lower down salary level.

As shown in fig 1: according to the model of specific factors, because of the inflow of labor to foreign-funded department, $VMP`_{Lf}$ moves leftward to $VMP``_{Lf}$. The balance point at this time is F`. $W`_{Lf}$ is the balance salary of two departments. The labor released from domestic department is completely absorbed by foreign-funded department. But, the fact is that foreign-funded department only absorbs those skilled labors and the labor used is L_f; on the other aspect, the outflow of technical talents of domestic department results in lowered capital productivity. VMP_{Ld} moves rightward to $VMP`_{Ld}$ and joins $VMP``_{Lf}$ at F``. Salary declines to W_d. Labor used is L_d. The labor from L_f to L_d constitutes structural unemployment labor, adding to unfairness. Therefore, FDI not only makes the income of factor owners different, but also affects the income of workers from different departments.

Conclusion and Enlightenment

1. The relation between foreign capital as specific factor and unfair income has been validated. Within a short term, FDI will really result in unfairness of income distribution. With the improvement of domestic technical level and decreasing gap of technology owned by foreign enterprises, this kind of income difference will be reduced. When some kind of new technology is invented, which has monopolistic advantage, it may cause income difference. Therefore, the influence of specific factors (including technical capital) on income distribution should fluctuate. This is the matter to be studied further.
2. Foreign capital is only one of the factors for income difference. The heterogeneity of labor is also another cause for enlarged income gap. Besides, some unreasonable and illegal factors (like smuggling and escape with capital) also contribute to income difference. The defects of coexistence of new and old systems after reform and opening (like dual exchange rate) are also the cause for enlarged income gap.

According to the above conclusion, we can get the following enlightenment

Firstly, technical difference is the core difference between foreign capital and domestic capital, so we need to increase FDI technical spillovers. In China, the FDI technical spillovers are mainly embodied by human capital and affect the part of soft technology with workers as carrier. The job-hutting of trained employees is an important approach for technical diffusion. Domestic enterprises should make great efforts in research and development, and work with foreign enterprises to set up R & D center to improve technical level.

Secondly, establish a perfect education system, improve the training and accumulation of human capital, upgrade the quality of people, and try to eliminate structural unemployment.

Finally, make related policies to encourage reasonable flow of labor. Refer to experiences

of developed countries, establish a perfect social security system to avoid poverty from getting worse in a short term, and keep income difference from enlargement.

References

1. Gene H.Chang "The cause and cure of China`s widening income disparity" China Economic Review 13(2005) 335-340.
2. Tao Yang "What has caused regional inequality in China" China Economic Review 13(2002) 331-334.
3. Hu Zhaoling "Economic globalization and income inequality"[J] World Economic 2006 [10].
4. Zhang Haoguang,Jiang Xiulan "Have to stop the expanding of China`s income gap which is caused by foreign investment"[J] Finance&Economics 2006 (2).
5. Fan Huiyan ,Duan Junshan "FDI and China`s income distribution[J] Finance&Economics 2005 (2).
6. Zhai Yinyan ,Sun Yan "The influence of technology and international trade on income distribution" [J] Finance theoriese and practice 2007 (10).

MANAGEMENT REQUIREMENTS FOR BUSINESS INTELLIGENT PROJECT DEVELOPMENT

Kamelia Stefanova, Dorina Kabakchieva

Abstract. Business Intelligence (BI) is a set of concepts, methods and processes that support the decision making process improvement, using different sources of information, applying previous experience and defining assumptions to obtain an accurate picture of the business dynamics. The logical architecture of BIS is structured in three main layers, Management of Data, Management of Data Integration and Analytical Models, and Management of Data Visualization, each of them providing specific tools for user interaction. Successful implementation of BIS requires appropriate use of all the data generated in the enterprise to empower the managerial capacity for transforming data into information and information into knowledge and competitive advantage. Strategic aspects at various levels, Information, Users and Organization, should be taken into consideration in BI project planning and development, in order those initiatives to be successfully fulfilled.

Keywords: Business Intelligence, Business Intelligent Systems, Business Intelligent Project Management.

1 Introduction

Companies today, in the era of the intelligent economy, are in strong competition that could be measured and analyzed by the level of successful management.

Corporate performance is challenged to meet the requirements of the dynamically changing market conditions and has to be based on proper and timely management decision making.

Organizations have to become intelligent in collecting, understanding and using data for making management decisions in order to successfully carry out business activities.

Business Intelligent Systems (BIS) become the key functional component of modern companies that links data, information, people and technology, and results in successful organization management.

Companies that follow the market trends, implement BI solutions to turn operational data into information and information into competitive advantages.

BISs satisfy the increasingly urgent demands for systems that can derive business value from big, diverse data streams (including machine data) and apply technologies such as predictive analytics and automated decision management.

2 Business Intelligence Concept

Business intelligence is a term that has been introduced in order to describe the new level of information systems growth - advanced analytics.

It encompasses different elements - technologies, applications, and processes for gathering, storing, accessing, and analyzing data to support management users taking appropriate decisions.

- The term "Business Intelligence" escalated mostly in the 1990s, but actually it has originated much earlier:
- First information systems that address the information support of the decision process are called Management Information Systems and emerged in 1970s. Their main characteristics are - Records and Reports that are static, two-dimensional and without analytic functions.
- The development of Information Systems in early 1980sis related to the Strategic Management (Executive Information Systems), assisting senior management of organizations. These systems are able to process Dynamic multidimensional queries and reports, prediction and forecasting, trend analysis, detailed research data, analysis of state and critical success factors.
- Mid1990s proved the BIS to be well introduced in the companies and numerous commercial software products elaborated and incorporated advanced analytical tools.
- Late 1990s gave the great push for additional new features development in the newly emerged business intelligence systems. That is the period when the new common understanding is established - all the information required by the organization management to support making timely and appropriate management decisions can be found in a business intelligent system covering the whole enterprise.
- After 2005 business intelligent systems begin to provide powerful analytic tools and features that serve innovative and advanced analytical tasks.
- Today, new generation of Business Intelligence is flourishing and expands in different development directions.

During the period of BI and BIS industry implementations many different definitions of related terminology can be figured out, refined and supplemented.

3 Business Intelligent Systems Architecture

BIS architecture is structured in three main layers with specific tools for each [1]:

- Management of Data – this layer includes relational databases and all enterprise data sources. The role of this layer is to collect, integrate and organize data from internal and external sources. Data sources contain historical and current data. According to the type and scope, enterprises could decide what alternative procedures for data integration could be selected.
- Management of Data integration and Analytical Models – this layer is responsible for incorporating the enterprise data ina Data Warehouse, mainly implementing procedures for extraction, transformation and load (ETL).This layer figures out the

different type of analytical models to be used for data processing– analysis, forecasting, trends, patterns,etc.The main solutions implemented at this layer are: On-line Analytical Processing (OLAP), Data Mining, AnalyticalReporting.

- Management of Data Visualization – this layer's main tasks include a visual drill capacity to identify complex relationships and sophisticated graphical instruments and tools for reporting and presenting data in the most convincing and friendly manner. This layer could also incorporate a business intelligence portalto integrate data and information from a range of applications and repositories to allow visualization through a unique Web interface [3].

The model of the BIS architecture, based on the three above described layers is presented on Figure 1.

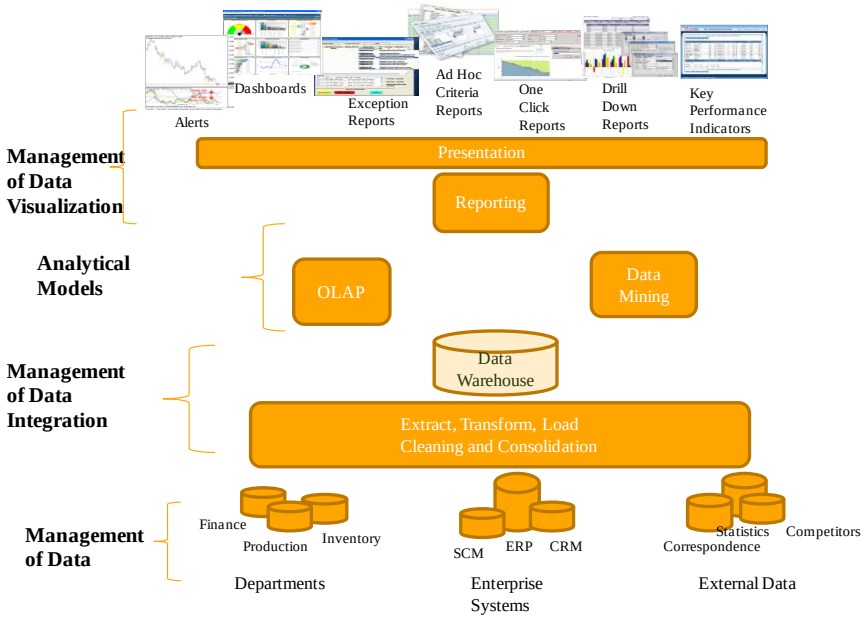


Fig. 1. A BIS Architectural Model

Modeling process typically follows an open approach whereby the model builders use their knowledge and intuition to construct and position the models.

4 Business Intelligent Project and Management

The information expansion is a problem faced by companies when they start organizing the data management and analytics [2]. Every day, executives make critical business decisions based on uncoordinated information from a variety of sources and they need to understand the process related to the BI project organization.

Information Strategy

- The BI strategy should take into account all sources of data used for business decisions: operational systems, departmental data marts, key spreadsheets, and unstructured data, Web pages, and other corporate systems.
- Data sources should include not only internal systems, but also information stored in the systems of customers, partners, and suppliers.
- The enterprise data warehouse is the central pillar of any BI strategy.
- Data of all diverse operational systems should be integrated, made consistent, and optimized for analysis.

Organizational Strategy

- The BI strategy should be in line with the organization as a whole and processes that require information to be managed to reach the corporate asset.
- A working group for the BI project management should be established to ensure the realization of the information value added.
- Appropriate governance should be introduced to set priorities and align the information use with the overall strategy of the organization.

User Strategy

- The most important approach to the BI project is that information stays irrelevant until users consider it for decisions in the way the business operates.
- The BI strategy should examine who needs which information, from which systems, and how to be interacted.
- The goal is to ensure that all people and processes have the information they need, when they need it, in order to fulfill their business roles.

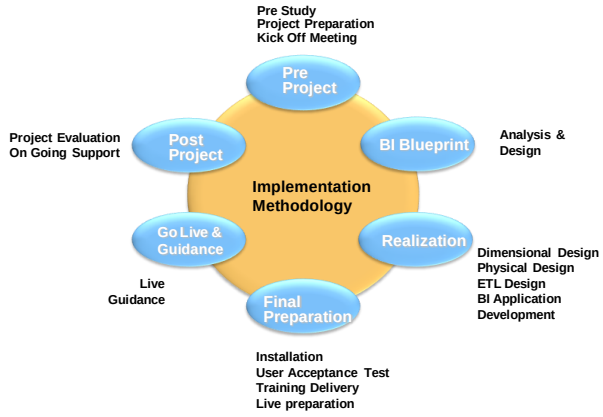


Fig. 2. BI Project Management Map

The phases within the BI Project Management Map and their main objectives are presented in Fig. 2.

5 Future Directions in BI Project Management

In-memory computing changes the terrain for BI and analytics

- BI and analytics systems are enabling greater use of expanded main memory. This is a cost-effective performance improvement because memory continues to get cheaper – not running analytic processes iteratively against the data.
- In-memory computing may not be the best option when data volumes are too large (compression technology is critical for in-memory systems) or data must be updated frequently.
- In 2013, in-memory computing will continue to shift the balance for BI and analytics systems.

Data visualization becomes essential as analytics follow the mainstream

- Data visualization and visual data analysis enable nontechnical users to see data patterns and trends they would have struggled to grasp with tabular reports, spreadsheets, and primitive graphics.
- More innovation in 2013 to support exciting and sophisticated graphics for BI and analytics, which will include more animation and intuitive, "gesture-based" data interaction.

Right Data at the Right Time

- With demand continuing to rise for more information and new ways of accessing, analyzing, and sharing it, nobody gets to rest in this industry.
- With sharper data insights, decision makers will have clear and comprehensive views of subjects of interest supported by the right data at the right time.

6 Conclusions

BI systems are constantly being developed and innovated to meet the ever-growing demands of business users, providing opportunities for real-time analysis of business data, and particularly the front-line operations. The implementation of advanced analytical methods and powerful visualization and alerting tools have led to the creation of dashboards supporting the decision making processes at all management levels – strategic, tactical and operational.

References

1. Bara, A., Botha, I. Diaconita, V., Lungu, I., Vekicanu, A., Velicanu, M. (2009). A model for Business Intelligence Systems' Development. Informatica Economică vol. 13, no.4/2009.
2. The Data Warehousing Institute (TDWI) (2012). Business Intelligence Fundamentals: from Data Warehousing to Business Impact.
3. Diaconita,V., Botha, I, Bâra,A., Lungu,I., Velicanu,M. (2010)."Two Integration Flavors in Public Institutions", Transactions on Information Science and Applications, May 2010.
4. Wixom, B., Watson, H. (2011). The BI-Based Organization.International Journal of Business Intelligence Research, Volume 1, Issue 1, January-March 2011.

Applying Data Mining Methods and Techniques for Student Data Analysis

Dorina Kabakchieva, Kamelia Stefanova

University of National and World Economy, Sofia, Bulgaria
dkabakchieva@unwe.bg, kstefanova@unwe.bg

Abstract. The paper presents the results of a study performed within a research project conducted at the University of National and World Economy. The main purpose of the study is to find out how the application of various data mining methods and techniques on the available university data could contribute to supporting the decision making processes concerning effective management, organizing successful marketing campaigns, improving teaching quality, analyzing student performance and applying individualized approaches to peculiar groups of students with similar characteristics.

Keywords: Educational Data Mining, Classification, Student Performance Prediction, Student Retention

1 Introduction

Universities today are operating in a very complex and highly competitive environment. They are collecting and keeping large volumes of data, referring to their students, the organization and management of the educational process, and other managerial issues. However, managerial decisions are rarely taken based on the information and knowledge that could be extracted from the analysis of the available data. In order to achieve this, they should be introducing innovative management approaches supported by progressive methods and techniques.

Modern universities should focus more on the profile of admitted students, get aware of the different types and specific student characteristics based on the received data and take a decision if they have all the data needed to analyze the students at the entry point of the university or they need other data to help the managers support their decisions in how to organize the marketing campaign and approach promising potential students.

"Educational Data Mining" (EDM) is a new stream in the Data Mining research field, focusing on the implementation of data mining methods and tools for analyzing data available at educational institutions. There are already a large number of research papers discussing various problems within the higher education sector and providing examples for their successful solutions reached by using data mining.

There are two papers available, in which extensive literature reviews of the EDM research field are provided. The first paper is published in 2007 and covers the research efforts in the area between 1995 and 2005 [1], and the second paper is provided in 2009 and refers to the research efforts during the period after 2005 [2]. Predicting student performance, retention of students, targeted marketing, enrollment management, and improving institutional effectiveness, are problems that are very often attracting the attention of

researchers and becoming the reasons for initiating data mining projects at higher education institutions.

This paper is organized in four sections. The reasons for the initiation of the performed research work are described in the Introduction section. The state-of-the-art in the Educational Data Mining research area, referring to predicting student performance, is briefly presented in the second section. A description of the data mining research project, initiated at one of the most prestigious Bulgarian universities, including the main goal, objectives and research questions, is also provided in that section. The applied research methodology is described in the third paper section. Some of the achieved results within the project are presented in the fourth section. The paper ends with Conclusions section, in which the main achievements are summarized, including interesting finding from the application of selected data mining methods. The next steps for future research are also defined.

2 Using Data Mining Methods for Predicting Student performance

The literature review reveals that predicting student performance has been of interest for various researchers during the last years. Kovačić uses data mining techniques (feature selection and classification trees) in [3] to explore the socio-demographic variables (age, gender, ethnicity, education, work status, and disability) and study environment (course programme and course block) that may influence persistence or dropout of students, identifying the most important factors for student success and developing a profile of the typical successful and unsuccessful students. Ramaswami et al. focus in [4] on developing predictive data mining model to identify the slow learners and study the influence of the dominant factors on their academic performance, using the popular CHAID decision tree algorithm. The implementation of predictive modeling for maximizing student recruitment and retention is presented also in the studies of Noel-Levitz in [5] and DeLong in [6]. Dekker et al. deal with predicting students drop out in [7]. The development of enrollment prediction models based on student admissions data by applying different data mining methods is the research focus of Nandeshwar in [8]. Yu et al. explore student retention in [9] by using classification trees, multivariate adaptive regression splines (MARS), and neural networks. Cortez and Silva attempt to predict student failure in [10] by applying and comparing four data mining algorithms - Decision Tree, Random Forest, Neural Network and Support Vector Machine. Vandamme et al. use decision trees, neural networks and linear discriminant analysis in [11] for the early identification of low, medium and high risk students.

The rationale behind the research work described in this paper is based on the great potential that is seen in using data mining methods and techniques for effective usage of university data, and the identified needs for supporting university management in decision making. Knowing the students better is expected to contribute to more efficient university enrollment campaigns, to attracting the most desirable students and to improving the organization and implementation of the educational and management processes at the university.

The research work is performed within a data mining research project, initiated at one of the most famous and prestigious Bulgarian universities. The project main goal is to reveal the high potential of data mining applications for finding information in the existing data to

support the university management in better knowing their students and performing more effective university marketing policy. The specific objective of the proposed research is to find out if there are any patterns in the available data that could be useful for predicting student performance, based on their personal, pre-university and university characteristic features.

3 The Applied Research Methodology

The university data mining project implementation is based on the CRISP-DM model (Cross-Industry Standard Process for Data Mining), a standard approach for data mining process realization. It is a cyclic approach, including six main phases – Business understanding, Data understanding, Data preparation, Modeling, Evaluation and Deployment, and a number of internal feedback loops between the phases, resulting from the very complex non-linear nature of the data mining process and ensuring the achievement of consistent and reliable results. The main project phases and activities, to be accomplished during the project implemented and based on the CRISP-DM model, are presented on Fig.1.

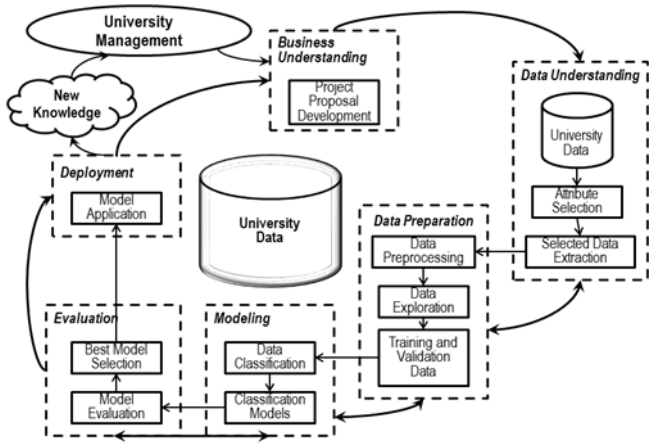


Fig. 2. University Data Mining Research Project Implementation

The application process for student enrollment at the University has been initially studied, including the formal procedures and application documents, in order to identify the types of data collected from the university applicants and stored in the university databases in electronic format. Discussions are also conducted with university management and representatives of the administrative staff, who are well informed and responsible for the university data collection, storage and maintenance. It is established that university data is basically stored in two databases. All the data related to the university admission campaigns is stored in the first database, including personal data of university applicants (names, addresses, secondary education scores, selected admission exams, etc.), data about the organization and performance of the admission exams, scores achieved by the applicants at the admission exams, data related to the final classification of applicants and student

admission, etc. All the data concerning student performance at the university is stored in the second database, including student personal and administrative data, the grades achieved at the exams of the different subjects, etc.

The defined business problem, predicting student performance based on their personal, pre-university and university characteristic features, in data mining terms is considered a classification problem to be solved. There are various data mining methods and techniques available for developing models for prediction, including rule classifiers, decision tree algorithms, regression techniques, Bayesian methods, neural networks, etc. These methods are based on different algorithms and follow different procedures for generating classification models, and all of them have advantages and drawbacks. Four classification algorithms are initially applied on the selected and preprocessed data – a rule learner OneR, a decision tree algorithm, a neural network and a k-Nearest Neighbour algorithm.

The OneR classifier is a rule learning algorithm generating a one-level decision tree expressed in the form of a set of rules that all test one particular attribute. It is a simple, cheap method that often produces good rules with high accuracy for characterizing the structure in data. This classifier is often used as a baseline for the comparison between the other classification models, and as an indicator of the predictive power of particular attributes.

Decision trees are powerful and popular tools for classification. A decision tree is a tree-like structure, which starts from root attributes, and ends with leaf nodes. Generally, a decision tree has several branches consisting of different attributes, the leaf node on each branch representing a class or a kind of class distribution. Decision tree algorithms describe the relationship among attributes, and the relative importance of attributes. The advantages of decision trees are that they represent rules which could easily be understood and interpreted by users, do not require complex data preparation, and perform well for numerical and categorical variables.

Neural networks produce classification models in the form of a mathematical model, consisting of interconnected computational elements (neurons) and processing information using a connectionist approach to computation. They are used to model complex relationships between inputs and outputs and very often yield very good results.

The k-Nearest Neighbor algorithm (k-NN) is a method for classifying objects based on closest training examples in the feature space. k-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The k-NN algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). The best choice of k depends upon the data; generally, larger values of k reduce the effect of noise on the classification, but make boundaries between classes less distinct. The accuracy of the k-NN algorithm can be severely degraded by the presence of noisy or irrelevant features, or if the feature scales are not consistent with their importance.

The software tool that is used for the project implementation is the open source data mining software packages WEKA.

4 The Achieved Results

The provided initial dataset contained data about 10330 students that have been enrolled as university students during the period between 2007 and 2009, described by 20 parameters, including gender, birth year, birth place, living place and country, type of previous education, profile and place of previous education, total score from previous education, university admittance exam and achieved score, total university score at the end of the first year, etc. The data is carefully studied and subjected to many transformations.

The final dataset, on which the selected classification data mining algorithms are applied, contains 10067 instances and 14 attributes, both categorical and numerical variables. Thirteen of the attributes are used as input variables, including three types of data about the university students – personal, pre-university and university data. Personal data includes *gender*, *age* and *birth year*. Pre-university data includes *secondary school profile (e.g.* languages, economics, technical, arts, sports, etc.*), secondary school location*, *secondary education scores*, *successfully passed admission exam* (Literature, Mathematics, History, Geography, Economics), *successfully passed admission exam score*, *total score for admission/rejection*, etc. University data includes university specialization, number of failures during the first year of education, etc. The fourteenth attribute is used as a predicted variable (the class) – it is a binary categorical target variable, constructed based on the original numeric parameter "University average score" (the average numeric score achieved by the students at the end of the first year at the University). The predicted variable has two distinct values, corresponding to the two classes in which the students are classified – Weak and Strong. Since a six-level scale is used in the Bulgarian educational system for evaluation of student performance at schools and universities, the students with average university score that is lower than 4.50 are classified as "Weak", and the students with average university score equal or higher than 4.50 are classified as "Strong".

The selected data mining algorithms are applied to the dataset using the holdout method (WEKA "Percentage Split" test option, 66%/34%). The dataset is divided into 3 parts and, each time an algorithm is run, 2/3 of the data is used for training of the classification model and 1/3 of the data is used for testing and evaluation of the model.

The four WEKA data mining classification algorithms that are applied on the available data include: the rule classifier 1R (OneR), the decision tree algorithm J48 (J48), the neural network Multilayer Perceptron (NNet) and the kNN algorithm IBk (kNN).

The results from the evaluation of the classification models, generated with the selected four data mining algorithms, are presented on Fig.2. The models are compared according to the achieved Accuracy/Error of Prediction, referred as the % of correctly/incorrectly classified instances.

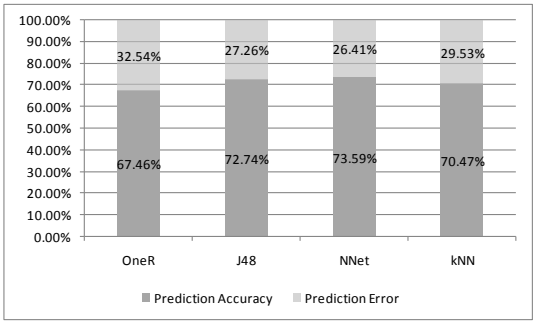


Fig. 2. Results from the data mining classification models comparison

The results reveal that the highest classification accuracy is achieved for the Neural Network model (73.59%). However, the disadvantages of that model are its complexity and the difficult understanding and interpretation by users. That model is suitable to use when the main objective is to predict the students' class (Weak or Strong), but there is no need to provide the exact reasons for the distribution.

The Decision Tree model is also working with a high accuracy of prediction (72.74%). The advantages of this model are that it is easily interpretable because it produces a set of understandable rules, and that it is working well with both, nominal and numeric variables.

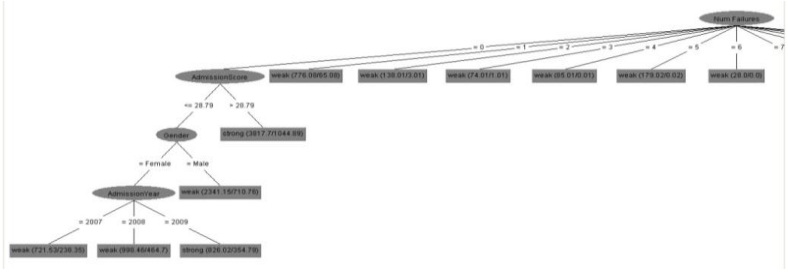


Fig. 3. The Decision tree model generated by applying WEKA J48 algorithm on the data

The decision tree model, that is achieved for M=500 (M is the minimum number of instances per leaf), is presented on Fig.3.

Easily interpretable rules can be derived from the decision tree models as follows:

IF NumFailures=1,2,3,4,5,6,7,8,9,10,11,12, THEN Class=Weak
IF NumFailures=0 AND AdmissionScore>28.79, THEN Class=Strong
IF NumFailures=0 AND AdmissionScore≤28.79 AND Gender=Male, THEN Class=Weak
IF NumFailures=0 AND AdmissionScore≤28.79 AND Gender=Female AND AdmissionYear=2009, THEN Class=Strong
IF NumFailures=0 AND AdmissionScore≤28.79 AND Gender=Female AND AdmissionYear=2007,2008, THEN Class=Weak

The rules can provide clear explanation about the characteristic features of the students distributed in the two classes – Weak and Strong. The attributes that appear at the

upper levels of the tree include *Number of Failures*, *Admission Score* and *Gender*, and these are the features that seem most important for the classification of the students by this model.

The kNN model (70.47%) and the OneR rule classifier (67.46%) are predicting the student class with lower accuracy. The OneR algorithm uses the "Admission Score" attribute for the classification which once again proves that the Admission Score parameter is very informative for recognizing strong and weak students.

Conclusions

The classification models, generated by applying the selected four data mining algorithms – OneR Rule Learner, Decision Tree, Neural Network and K-Nearest Neighbour, on the available and carefully pre-processed student data, reveal classification accuracy between 67.46% and 73.59%. The highest accuracy is achieved for the Neural Network model (73.59%), followed by the Decision Tree model (72.74%) and the k-NN model (70.49%). The data attributes related to the students' University Admission Score and Number of Failures at the first-year university exams are among the factors influencing most the classification process.

The results and conclusions of the performed study will be used for defining the further steps and directions for the university data mining project implementation, including possible transformations of the dataset, tuning the classification algorithms' parameters, etc., in order to achieve higher accuracy for the classes' prediction. Recommendations will also be provided to the university management, concerning the sufficiency and availability of university data, and related to the improvement of the data collection process.

References

1. Romero, C., Ventura, S. (2007). Educational Data Mining: A Survey from 1995 to 2005. Expert Systems with Applications 33, 2007, pp.135-146.
2. Baker, R., Yacef, K. (2009). The State of Educational Data mining in 2009: A Review and Future Visions. Journal of Educational Data Mining, Vol.1, Issue 1, Oct. 2009, pp.3-17.
3. Kovačić, Z. (2010). Early Prediction of Student Success: Mining Students Enrolment Data. Proceedings of Informing Science & IT Education Conference (InSITE) 2010, pp.647-665.
4. Ramaswami, M., Bhaskaran, R. (2010). A CHAID Based Performance Prediction Model in Educational Data Mining. IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 1, No.1, January 2010, pp.10-18.
5. Noel-Levitz White Paper (2008). Qualifying Enrollment Success: Maximizing Student Recruitment and Retention Through Predictive Modeling. Noel-Levitz, Inc., 2008.
6. DeLong, C., Radclie, P., Gorny, L. (2007). Recruiting for retention: Using data mining and machine learning to leverage the admissions process for improved freshman retention. Proc. of the Nat. Symposium on Student Retention 2007.
7. Dekker, G., Pechenizkiy, M., Vleeshouwers, J. (2009). Predicting Students Drop Out: A Case Study. Conference Proceedings of the 2nd International Conference on Educational Data Mining (EDM'09), 1-3 July 2009, Cordoba, Spain, pp.41-50.
8. Nandeshwar, A., Chaudhari, S. (2009). Enrollment Prediction Models Using Data Mining. Available at: http://nandeshwar.info/wp-content/uploads/2008/11/DMWVU_Project.pdf

9. Yu, C., DiGangi, S., Jannasch-Pennell, A., Kaprolet, C. (2010). A Data Mining Approach for Identifying Predictors of Student Retention from Sophomore to Junior Year. Journal of Data Science 8(2010), pp.307-325.
10. Cortez, P., Silva, A. (2008). Using Data Mining to Predict Secondary School Student Performance.EUROSIS, A. Brito and J. Teixeira (Eds.), 2008, pp.5-12.
11. Vandamme, J., Meskens, N., Superby, J. (2007). Predicting Academic Performance by Data Mining Methods.Education Economics, 15(4), pp.405-419.

E-government in the public sector - challenges and paradoxes

Rosen Kirilov

Abstract. IT professionals and ordinary users are witnessing the continuing efforts of state and local government to achieve an electronic management. At the same time researches made and general impressions show us clearly, that society does not use the full potential of electronic services, provided by the administrations. The paper analyzes the money invested for the electronization of the public sector and the results obtained. On this basis conclusions are formulated.

Keywords: information technologies, e-management, public sector.

This paper is intended to provide support for answering the question "is e-governance possible in the public sector?". Implication that we use is "when will be realized the true e-governance in municipalities? Is it possible? What are the main problems and challenges? ". Since the answers to these questions in the context of the concept of e-government are quite brave, we aim to adduce evidence in support of our view that urgent action is needed to introduce a uniform computer information system in municipalities. Whether this view is correct or not time will tell, but at this stage will cite evidence to support this claim.

In early 2012 a team of professionals in local government and information technologies realized a survey on the level of information technologies in 55 municipalities in Bulgaria. The survey was conducted by sending a questionnaire by mail.

Here are some of the results that we can cite in support of the thesis of the necessity of introducing a uniform computer information system in municipalities:

1. The question about how many different software applications are used in the surveyed municipality has the following answers:

- An average of 55 municipalities surveyed used 7.8 different software applications;
- The minimum number of applications is 2;
- Maximum number of software applications is 30;

The distribution of results is sufficiently indicative of the fact that the municipalities in Bulgaria continue the practice of using multiple software applications that are actually unrelated to each other. These results contradict our understanding that e-governance is possible only when database systems and software applications are integrated with each other.

2. The question “Does your municipality have software applications designed specifically for it?” makes it clear, that only 12 of the 55 municipalities surveyed have such applications. We think this percentage (about 22%) is too low. Here are a few facts to support this statement:

- The first step in implementing a data project is realized need of automation in an activity. Discussed in terms of the design of information systems, we believe that this step

(phase) is essential. In practice, in most cases, the awareness of the need for automation arises knee-jerk to the lower levels of management (operational and tactical) and subsequently is transmitted to higher levels (tactical and strategic). Reaching the desire for automation to strategic management level is essential, because it commits the company's senior management with the implementation of IT projects;

- Business processes across organizations and municipalities have a number of peculiarities. These features can be classified as specifics of interpretation of laws and regulations, refracted through the prism of applied best practices in local government. The high level of detail in the description and modeling of business processes can lead to the identification of municipality typical steps of these processes. All this, along with many attempts to describe the practice of all processes within an organization, confirm the assertion that specific business processes require the creation of specific software tools;

- The organizational structures of the different municipalities are different, which was confirmed by the study. Different organizational structures make use of the same 'type' software difficult. This means that the roles of individual users in the system are defined once and are never allowed to be parameterized;

- We can indicate a number of other examples to support our thesis that creating a unified software tool and its practical use in a wide range of municipalities would be very difficult (at times tending to be impossible) and will lead to low efficiency of the software project.

We believe that there are some steps, which are necessary in order the e-governance to be possible (Fig. 1).

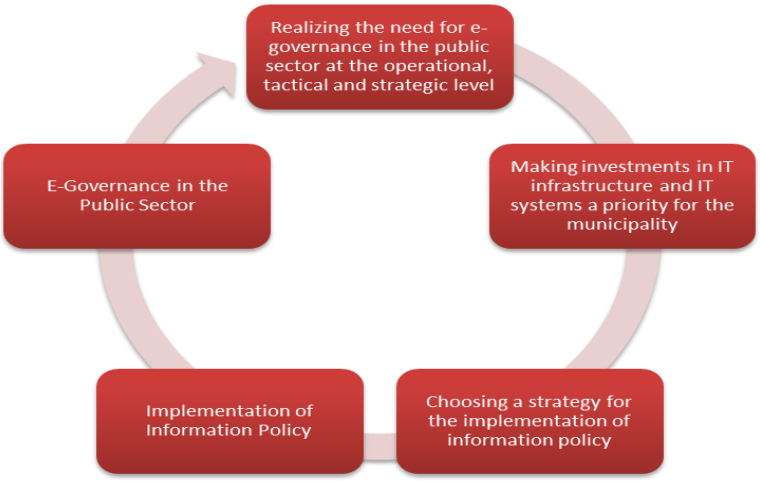


Fig. 1. Phases for the implementation of e-governance in the public sector

In terms of the goal of this paper it is necessary to address the question of the potential of local IT infrastructure. It can be estimated based on the study of factors that determine the adoption of modern computer systems.

The analysis of the local information infrastructure at national level in a sample of 55 municipalities identifies the following trends (Table 1).

Table 1. Structure of the municipalities analyzed by number of employees

Employees	Municipalities
Less than 50	5
51 to 100	22
101 to 200	16
More than 200	11

As seen from Table 1, most municipalities have indicated that the number of administrative employees is between 51 and 100 people. These are 22 of the 54 municipalities analyzed. Next are a number of municipalities with a number of employees between 101 and 200 – 16 municipalities. Third place is for municipalities that have over 201 employees – 11 municipalities. The smallest is the number of municipalities with number of employees in the municipal administration less than 50 people. The distribution is shown in Fig. 2.

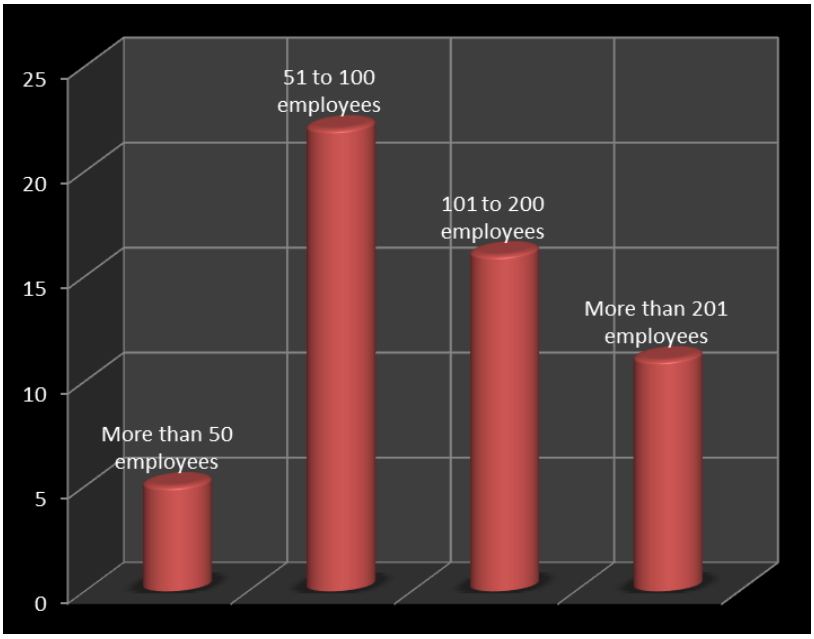


Fig. 2. Structure of total number of employees in their administration

From the perspective of this report it is interesting for us the structure of employees in information technologies (Table 2).

Table 2. Structure of the municipalities analyzed by number of employees in IT

Employees	Municipalities
Less than 50	51
51 to 100	2
101 to 200	0
More than 201	1

As seen from the data the largest number of municipalities surveyed have indicated that the number of IT staff is under 50 people – 51 municipalities. This gives grounds to believe that the ratio between the number of employees in information technologies and the number of employees in larger municipalities is quite unbalanced.

Graphical distribution of the results of the number of employees in information technologies is given at Fig. 3.

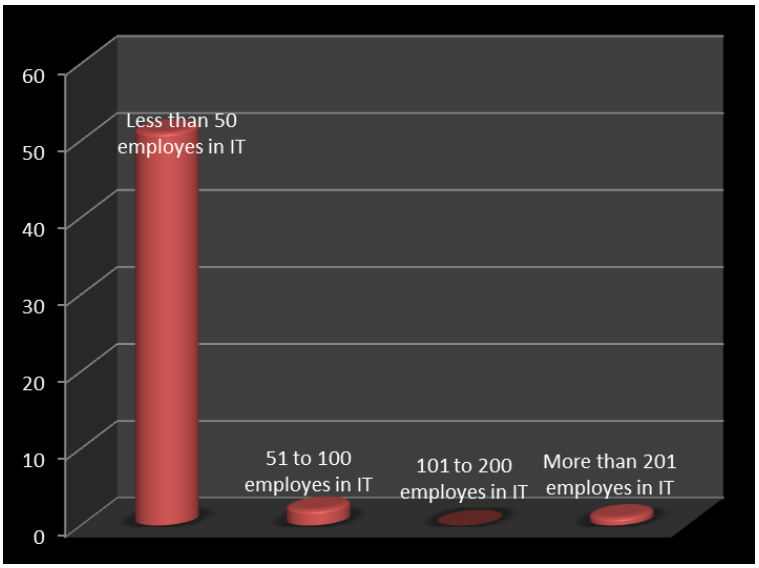


Fig. 3. Structure of total number of employees in information technologies in their administration

The results confirm the fact that as investments in information and communication technologies are not a priority for most municipalities in Bulgaria, the realization of e-governance is very difficult.

References

1. Kirilov, R., Information technologies in the financial system, Publisher "Avangard Prima", Sofia, 2007, ISBN 978-954-323-309-0.

Analysis of conflicts during the integration of heterogeneous multi-dimensional Data marts

Alexandrina Murdjeva[1], Geno Stefanov[1]

[1] Department of Information Technologies and Communications, Faculty of Applied Informatics and Statistics, University of National and World Economy, Hristo Botev str. , Studentski grad, 1700 Sofia, Bulgaria
{amurdjeva, genostefanov}@unwe.bg

Abstract. A company Data Warehouse is usually built following a bottom-up approach, where the Data Warehouse is a set of materialized views over the whole information system of the organization, namely a Data Mart. The problem of integrating autonomously developed data marts can arise when an organization needs to merge data residing in independently developed data warehouses in order to make a better decision. Data mart architecture is based on three basic elements: the facts to be analyzed, the coordinates by which to analyze, and measures that allow a fact to be evaluated quantitatively. As a result of the Data mart structure there are two levels of heterogeneities – dimension and fact. The former depends on differences between both dimensions hierarchies, members of dimensions, and names of members, levels and dimensions. The latter kind of conflicts occurs when facts in different Data marts are in different names, values (inconsistent measures), formats or even different scale. This paper summarizes and analyzes the different types of conflicts and heterogeneities which can arise when independently developed data marts are being integrated. The first section of the paper presents the literature overview on the topic. Then the identified conflicts are reviewed and analyzed.

Keywords: Data warehouse, data marts, integration conflicts, data integration.

1 Introduction

In literature, a data warehouse (DW) can be defined as a highly specialized database system optimized for decision support. A company DW is usually built by consolidating the data in all subject areas that are meaningful and helpful for decision making, dividing them into different data marts (DM). Each DM provides a dimensional view of a single business process. Business Intelligence tools provide analysis functionality that plays the role of query interface between the users and the DMs [1].

One problem that needs to be solved in many practical cases is the integration of data marts that have been developed and operated independently. In our paper we are trying to set the drivers for such integration. In many cases, data which resides in multiple and independently developed data sources is needed for decision-making. For example, if we want to compare several KPI from different DM or to define new KPI, which is defined by the KPIs residing in several different DMs, we face two possible choices – building a new DM or the integration of the existing ones. Building a new DM is a costly and time consuming task, therefore it would be better if we have mechanisms for integration of the

existing DMs. Another case where DM integration is needed is in mergers and acquisitions of different companies. In this case for example one company acquires another company and the DMs in the acquired company should be integrated into the DW of the acquiring company. Here we have the same possible choices as in the previous case.

The multidimensional model (MDM) on which the concept of DW and DM are based is built up of three basic constructive elements: the facts which are analyzed, the dimensions (coordinates of the fact) and the measures which allow for the quantitative evaluation of the facts [2]. Very often in large organizations the need for integrating independently developed DMs arises. These data marts should be based on common dimensions and facts, but in many cases different departments of one company develop their own data marts and their integration becomes a difficult task. The difficulties come out of the heterogeneities of the MDM elements and can be classified as dimension conflicts and measure conflicts from a semantic point of view. The dimension conflicts occur when the dimension schema structures, dimension members or the naming of semantically related dimensions have semantic discrepancies. The measure conflicts occur when the measures in different DMs are in different names, values (inconsistent measures), formats or even different scale. Some work has been done on the problem of resolving the conflicts occurring in data mart integration [3], [4].

Dimension conflicts can be further classified as dimension schema conflicts, dimension member conflicts and naming conflicts. Dimension schema conflicts occur when two dimensions have different hierarchies with different levels. Dimension member conflicts occur when two dimensions have mismatched members which correspond to the same level. Naming conflicts occur when two semantically related dimensions and their members or levels have different names.

Most often from a structural point of view the possible conflicts are dismissed, because they are well studied in the literature. In the presented paper we have distinguished two major types of structural conflicts – Key conflicts and Value conflicts.

The goal of our paper is to summarize and analyze the conflicts during the integration of heterogeneous multi-dimensional Data marts.

The rest of the paper is organized as follows: Section 2 introduce the analysis and summarization of semantic conflicts occurring during the DM integration. In section 3 the structural conflicts are identified and analyzed. Finally in section 4, some conclusions and future work are inferred.

2 Semantic Conflicts

In recent years, the data integration community has been focusing on semantic heterogeneity of data. The semantic integration problem is far more complex than structural integration since semantic heterogeneity refers to how users interpret the meaning of given schema elements according to their understanding of the real world [5]. In the literature the following semantic conflicts during the integration of autonomously developed and operated Data Marts are considered [3], [4], [6] (fig. 1.).

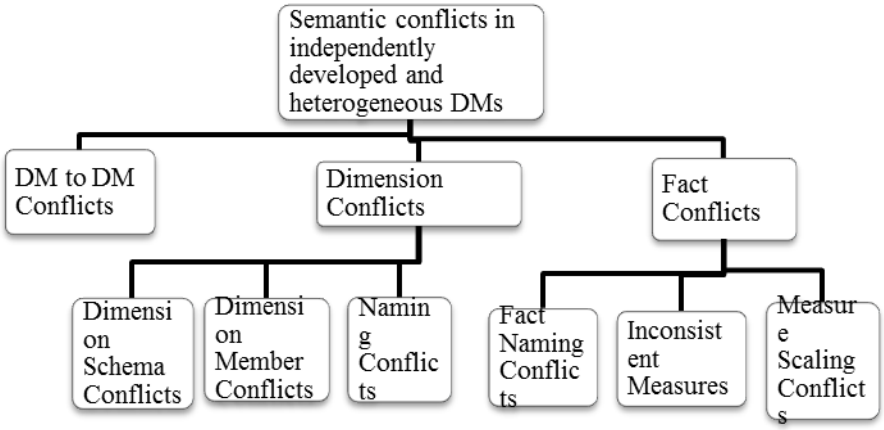


Fig. 1. Semantic conflicts in heterogeneous Data Marts

2.1 Dimensional Conflicts

Dimension conflicts can be classified as dimension schema conflicts, dimension member conflicts and naming conflicts as shown in fig. 1. These conflicts are result from the heterogeneity between the integrated dimensions.

Dimension schema conflicts occur when two DMs have different dimension hierarchy, with possibly different dimension levels. There are several degrees of *correspondence* among sets of levels depending on what is the equivalence of the levels and their hierarchy [6]. Two levels are equivalent if they model the same ontological concept. We have recognized four types of conflicts – (1) Non-corresponding dimension schemas (2) Partial-corresponding dimension schemas (3) Flat corresponding dimension schemas and (4) Domain conflicts. Non-corresponding dimension schemas conflict occurs when there are no common aggregation levels. Partially corresponding dimension schemas conflicts occur when the level sets of two dimension schemas have at least one, but not all equivalent level(s) in common. Flat corresponding dimension schemas conflict occurs when two dimension schemas have the same levels, but a different hierarchy. Domain conflicts occur when two attributes in level schemas of autonomous dimensions assign different sets of allowed values.

Heterogeneous members among dimensions can only be recognized correctly between homogeneous dimension schemas. This means that the dimension schema conflicts should be reconciled before the dimension member conflicts. Dimension member conflicts occur when two dimensions have mismatched members. Conflicts among dimension members may be caused either by single values of member attributes or by the member extensions as a set. Therefore, we further categorize (1) heterogeneities across single members in the dimension extensions, and (2) heterogeneities among the among the member sets as a whole. In order to determine and classify the dimension member conflicts we have to define member equivalence. Two dimension members are equivalent if they describe the same real world

entity. If two dimension members are not equivalent then the integration of the dimensions is not possible. If two dimension members are equivalent then the following conflicts can occur: heterogeneous roll-up functions and overlapping member extensions. The former conflicts occur when single member value rolls up to ambiguous parent member value. The letter conflicts occur when sets of members among two equivalent levels contain at least one pair of equivalent members.

Naming conflicts among dimensions can be defined as the occurrence of different words as labels of semantically equivalent entities in multi-dimensional Data Marts across schema boundaries. Due to the lack of standardized ontological vocabularies for the Data Warehousing domain, every organization usually maintains its own "private" vocabulary, which is well-known to the members of the organization but not to the public. Thus, it is common that heterogeneous vocabularies underlie the models of semantically "same" ontological concepts in the schema of a Data Mart. The naming conflicts across dimensions become obvious as soon as independently developed, but semantically related multi-dimensional schemas have to be integrated. In general, naming conflicts in the dimensional model of a data mart can be either homonyms or synonyms. Homonyms to attribute or level names occur if an identical name matches a different concept. In contrast, synonyms occur if several different level or attribute names map to the same concept. In the multi-dimensional model, heterogeneous attribute names affect dimension schemas and fact schemas alike. Naming conflicts among autonomous dimensions may occur within: (1) the name of the dimension, (2) level names within hierarchies, and (3) level attribute names in level schemas.

2.2 Fact Conflicts

Fact conflicts occur when the measures in different fact schema are in different names, different values (inconsistent measures), different formats, or even different units. Such conflicts can be further classified into three categories: Naming conflicts: such conflicts occur when two local fact schemas have semantically related measures and dimensional attributes with mismatched names. Inconsistent measures: such conflicts occur when two local fact schemas have semantically related measures with mismatched values. Measure scaling conflicts: such conflicts occur when two local fact schemas have semantically related measures with mismatched scales.

Fact naming conflicts can affect the following elements of fact schemas:

1. The name of the fact table;
2. Dimensional attributes of the fact schema;
3. Measures names;

2.3. Data Mart to Data Mart Conflicts

These conflicts occur when semantically related local Data Marts are created by different dimensional models. For example, suppose a Data Mart DM1 was modeled in star schema, but the other Data Mart DM2 was constructed via snowflake. Generally, a Data Mart allows the human business analyst to interpret numerical benchmarks modeled as measure variables in the particular context or business perspective given by the DM's dimensions. Sometimes, part of the fact context is modeled using elements at the schema

once and elements at the instance level once across distinct autonomous Data Marts. In particular, the multi-dimensional model provides measure attributes and dimension members for representing context information within the schema or instance of Data Mart, respectively. If dimension members are used, the context information is directly and conveniently visible from the DM as an additional dimension variable in the fact schema. Otherwise, special names for measures in the cells encode the context information implicitly. In the latter case, the fact context is only obvious from the presence of several measure variables, the semantics of which are closely related.

In most cases it is adequate to represent the context of facts exclusively within the members of dimensions. The reason for this is twofold. First, the names of measure variable subsets representing some part of the fact context can be viewed as enumeration of the values of an implicit, invisible context dimension. Second, the dimension schema provides additional properties e.g., hierarchies, non-dimensional attributes that allow the members to be modeled more precisely. Thus, using dimension members for expressing the fact context leads to more concise and more easily comprehensible DM's schema.

3 Structural Conflicts

Structural conflicts are well analyzed in the literature on the domain of database integration, but not as well analyzed when it comes to the integration of multi-dimensional data models. In our paper we are trying to analyze and classify them. In [7] the authors propose a classification of conflicts occurring in databases. They distinguish the following conflicts:

Value-to-Value conflicts - these conflicts occur when databases use different representations for the same data.

Value-to-Attribute conflicts - these conflicts occur when the same information is expressed as values in one database and as (an) attribute(s) in another database.

Value-to-Table conflicts - these conflicts occur when the attribute values in one database are expressed as tables in another database.

Attribute-to-Attribute conflicts - these conflicts are caused by using different definitions for the semantically equivalent attributes in different databases.

Attribute-to-Table conflicts - these conflicts occur if an attribute of a database is represented as a table in another database.

Table-to-Table conflicts - these conflicts are caused by representing the information of a set of semantically equivalent tables in a different number of tables in other databases.

Based on this classification we can distinguish the following (fig. 1.) structural conflicts in autonomously developed and possibly heterogeneous Data Marts.

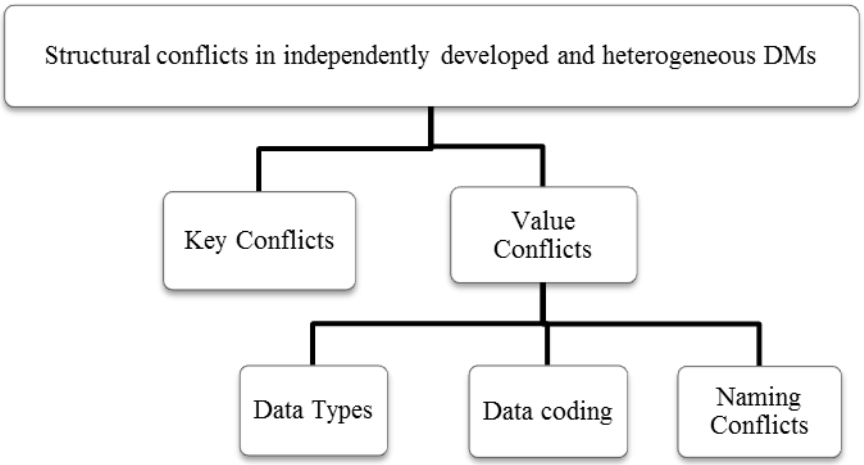


Fig. 2. Structural conflicts in independently developed and heterogeneous DMs

According to our classification there are two types of structural conflicts: Key conflicts and Value conflicts.

Key conflicts may occur in independently developed DMs when a different set of attributes are used for the unique identification of any tuple either in dimension or fact schema.

Value conflicts can be further classified as data type, data coding or naming conflicts. The data type conflict occurs when different data types are used for storing semantically related data. The data coding conflict can occur when different methods for coding the data are used. The naming conflicts are similar to those described previously.

4 Conclusions and Future Work

In this paper we have focused on the problem of integration of Data Marts, with the main emphasis on classifying and analyzing different types of conflicts.

We have classified the conflicts as either semantic or structural. The semantic integration problem is far more complex than structural integration since semantic heterogeneity refers to how users interpret the meaning of given schema elements according to their understanding of the real world. Thus, we propose that the order of conflicts resolution should be (1) resolving semantic conflicts and then (2) structural conflicts when integrating Data Marts autonomously. In general, heterogeneities among dimensions are repaired before conflicts among facts, because fact schemas logically depend on the dimension schemas in the multi-dimensional data model.

References

1. Kimball, R., Ross, M.: The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (2nd Ed.), John Wiley & Sons (2002)
2. Vassiliadis, P. and Sellis, T. K.: A survey of logical models for OLAP databases. (1999)
3. Tseng, F. S. C., Chen, C-W.: Integrating heterogeneous data warehouses using XML technologies. Journal of Information Science, 31(3):209-229 (2005)
4. Torlone, R.: Two approaches to the integration of heterogeneous data warehouses, Distrib. Parallel Databases 23(1):69-97 (2008)
5. Diamantini, C., Potena, D.: Data Mart Integration at Measure Level (2011)
6. Berger, S., Schrefl, M.: From federated databases to a federated data warehouse system. HICSS (2008)
7. Lee, C., Chen, C.-J., Lu, H.: An aspect of query optimization in multidatabase systems (extended abstract). p. 28-33 (1995)

Testing the Volume and Query Performance in Relational Databases in Case of Using Natural or Non-Natural Keys

Mitko Radoev

Department of Information Technologies and Communications, Faculty of Applied Informatics and Statistics, University of World and National Economy

Abstract. Choosing a primary key is of a particular importance in the design of a relational database. What are the possible keys, i.e. which columns or combinations of columns are candidate keys? Which of these candidate keys is best suited for a primary key?

For quite some time, a trend is gaining more popularity for choosing a primary key - not to use the existing columns of the relationship but to create new, additional column to ensure the uniqueness and to act as a primary key. Recently it became a common practice to use artificial or surrogate identifiers, even in cases where the relationships already have perfectly suited natural identifiers.

Using artificial keys is motivated mainly by practical considerations - the keys are shorter, do not have cascade update when changing the key, queries run faster etc. Does using of non-natural keys really reduce the size and increase the performance of the database? This study is testing what is happening with the database in adopting one or the other approach.

Keywords: primary key, natural key, non-natural key, query performance

1 Introduction

Over the years, almost from the inception of the relational data model has been discussion about the possibilities of using artificial keys and their advantages and disadvantages.

Opinions are divided on this issue. Even the world-recognized authorities in the field of databases have opposing views.

Recently it became a common practice to use artificial or surrogate identifiers such as:

- AutoNumber,
- Identity,
- Serial,
- Sequence, etc.

1.1 Definitions

Different authors use different terms such as: natural key, relational key, user-controlled key, artificial key, surrogate key, exposed physical locator etc. For the purposes of this study will distinguish two main categories of keys:

Natural key: A key composed of the attributes of the relationship, each of which has a value, which is defined outside the database. Otherwise, we will call natural these keys, which are based on the attributes of the relationship existing in the real world. It does not matter whether these attributes are physical characteristics or other intrinsic properties of objects or are given from state or other authority.

Non-natural key: A key that was created and exists only in the database, which has no analogue in the real world. Otherwise, a non-natural key will understand each key that is not natural, including artificial keys, surrogate keys and physical locators.

1.2 Related Work

For the first time the idea of using surrogate keys is set in 1976 by Hall, Owlett and Todd [1].

In 1979, during the presentation of the RM/T model Code adopts the idea by proposing the use of surrogate keys (surrogates), which are created by the DBMS and are unique to the entire database. Users will not be able to control the value of surrogate keys and this value will not be visible to them. Code, however, specifies that introduction of the surrogates does not make user-controlled keys obsolete. Users will often need entity identifiers (such as part serial numbers) that are totally under their control, although they are no longer compelled to invent a user-controlled key if they do not wish to [2].

Most works do not contain experimental data. In recent years, several experimental studies were presented, such as [3] and [4]; however they are limited in scope.

1.3 Objectives

The main objective pursued by the present study is to determine the impact of the choice of a primary key on the size of the database and the time of execution of queries.

This research will provide contribution to the following areas:

- Design of a hybrid schema in which some relations have natural and others non-natural keys;
- Create multiple queries for data retrieval (Select), and the modification (Insert, Update, Delete), corresponding to the actual queries used in an application.

2 Testing the Volume of the Database

Here are some of the preliminary results of the tests. In the *Table 1* you can see the volume of the database in three different cases:

- In case of schema with natural keys
- In case of schema with artificial keys
- In case of hybrid schema

Table 1. Size of the Database.

Parameter	Database with natural keys	Database with non-natural keys	Database with hybrid schema
Number of tables	17	19	19
Number of indexes	46	56	50
Size of the database	56,934,400	29,417,472	30,306,304
Percent	*100.00%*	*51.67%*	*53.23%*

As seen from the results, the size of the database with non-natural keys and hybrid schema is almost half the size of database with natural keys. Changing of the size of the tables can be calculated using the following formula:

$$\sum_{i=1}^{m} s_i.r_i - \sum_{j=1}^{n} (n_j - s_j).r_j \quad (1)$$

Where:

s is the size of the non-natural key,
n is the size of the natural key,
r is the number of the records,
m is the number of the tables with non-natural key,
n is the number of the related tables with distributed foreign key.

The same formula can be used to calculate the change in the indexes.

3 Testing the Query Performance

Tests with changing the data in the main tables are conducted to measure the performance of the database. Queries are performed to insert, update or delete 1000 records in each of these tables.

3.1 Testing Performance of Insert Queries

Table 2 presents the execution times of queries, which are inserting new records in the main tables of the database:

Table 2. Performance of Insert Queries (sec.).

Query name	Database with natural keys	Database with non-natural keys	Database with hybrid schema
InsertProperty	0.4688	0.5000	0.4844
InsertTenant	0.3438	0.4063	0.3906
InsertContract	0.5625	0.5469	0.5156
InsertCharge	0.6250	0.6094	0.6094
InsertPayment	0.7344	0.6563	0.6250
InsertPaymentDetail	0.4531	0.4844	0.5000
InsertAccount	0.5469	0.5469	0.5000
Average	*0.5335*	*0.5357*	*0.5179*

3.2 Testing Performance of Update Queries

Table 3 presents the execution times of queries for modification of the contents of the records in the main tables of the database.

Queries which are changing the contents of the primary key are presented separately.

Table 3. Performance of Update Queries (sec.).

Query name	Database with natural keys	Database with non-natural keys	Database with hybrid schema
UpdateProperty	0.4375	0.3750	0.3594
UpdatePropertyKey	1.2656	0.3438	0.3438
UpdateTenant	0.4219	0.3438	0.3594
UpdateTenantKey	1.1406	0.4219	1.0000
UpdateContract	0.4688	0.4063	0.3906
UpdateCharge	0.5156	0.5938	0.5781
UpdatePayment	0.7188	0.6406	0.6250
UpdateAccount	0.4844	0.4063	0.3906
Average	*0.6816*	*0.4414*	*0.5059*

3.3 Testing Performance of Delete Queries

Table 4 presents the execution times of queries, which are deleting records from the main tables of the database:

Table 4. Performance of Delete Queries (sec.).

Query name	Database with natural keys	Database with non-natural keys	Database with hybrid schema
DeleteAccount	0.5313	0.4688	0.4531
DeletePayment	41.8594	39.7500	39.2500
DeleteCharge	0.5781	0.5313	0.5000
DeleteContract	0.4844	0.4844	0.4375
DeleteTenant	0.5156	0.5781	0.5469
DeleteProperty	0.5625	0.6406	0.5938
Average	*7.4219*	*7.0755*	*6.9635*

4 Future Work

The study is still ongoing. Further tests must be performed:

- Testing Performance of Select Queries;
- Performing tests among different DBMS;
- Testing with a different number of records;
- Comprehensive evaluation of performance through the use of empirical data on the frequency of execution of queries.

Then the final conclusions will be made about the effect of using different keys on the performance of queries in the database.

References

1. Relations and Entities. Hall P., Owlett, J.,Todd, S. 1976. IFIP Working Conference on Modelling in Data Base Management Systems.
2. Codd, Edgar F. Extending the Database Relational Model to Capture More Meaning. ACM Transactions on Database Systems. Dec. 1979 г., Том 4, 4, стр. 397 - 434.
3. Link, S., Lukovic, I., Mogin, P. Performance evaluation of natural key and surrogate. Wellington, New Zealand : Technical Report ECSTR10-06, Victoria University, 2010.
4. Slavica Aleksic, Milan Celikovic, Sebastian Link, Ivan Lukovic, Pavle Mogin. Faceoff: Surrogate vs. Natural Keys. Advances in Databases and Information Systems: 14th East European Conference, ADBIS 2010, Novi Sad, Serbia, September 20-24, 2010, Proceedings. s.l. : Springer Berlin Heidelberg, 2010.

OPPORTUNITIES TO INTEGRATE TECHNOLOGICALLY INCOMPATIBLE BUSINESS APPLICATIONS

MONIKA TSANEVA, SM. KOUZMANOV

Department of Information Technologies and Communications, Faculty of Applied Informatics and Statistics, University of National and World Economy

Abstract. The structure of the software that is developed over the last decade has been designed with the idea of interoperability, i.e. modern software is composed of services - software components that isolate the implementation of certain functionality can be used in isolation or as part of other components (e.g. to be orchestrated) and introduce themselves via a standardized interface. The can be different types of program components that implement various activities of a business application, depending on the architecture of the program execution environment. Some types of software components are designed to be interoperable, other types are designed to run standalone and are "closed" for communication with the external software world. One option to create an integrated solution to information problems is a complete rewrite of all the "old" (welded) applications using a popular information technology system for building service-oriented (SOA). Apparently, this approach would be extremely costly and lengthy so that most companies can't afford it. This paper proposes an alternative one, by which the "old" applications or at least part of their functionality can be "packed" as a service, following already accepted standards in order to be usable by other systems. A concept of a software layer for existing business applications integration is proposed.

Keywords: Integration, service-oriented architecture, integration methodology, interoperability

Introduction

The majority of organizations have built business applications needed for their information service in decades. During this long period of time they repeatedly changed both the organizational structure and business logic of the enterprise and the information technology for the automation of these activities. As a result, many organizations operate in a number of business applications, that on the one hand are inconsistent in their architecture and the other hand meet the needs of the business. They have the following key issues:

- The need for multiple data entry into various systems, which increases processing and labour consumption and the number of entry errors
- Duplication of similar processes in different applications
- High volume communication flows between different applications

The main advantage of the interoperability of business information systems is that the business process can begin and end anywhere in the world, regardless of the required types of hardware and software for its automation.

Interoperability of business processes increases manifold usability of the system from the perspective of the user, as it increases the productivity of work. It reduces the

number of human errors and the time to extract useful information from the system. The user focuses on solving the task, not the on the tools used to implement this solution.

Interoperability of software components with different architecture

Software components that implement the various activities of the business application can be of different types, depending on the architectures of the software development and the program execution environments:

- services
- stored procedures in the database that underpins the application
- software classes and objects specific to used IDE
- system functions of the operating system

The structure of the software that has been developed over the last decade has been designed with the idea of interoperability Modern software is composed of services - software components that isolate the implementation of certain functionality. They can be used in isolation or as part of other similar components (e.g. be orchestrated) and present themselves via a standardized interface.

Services can be used in many different ways by different types of "customers":

- intelligent applications that interact with one or more services cached information extracted from them and provide both interactivity and support for batch processing of data
- intelligent devices e.g. phones
- user interface modules, e.g. portal applications or electronic controls of documents
- automation systems - customers that do not provide user interfaces, except in case of exception processing
- service orchestration processes, i.e. services that activate other services to implement aggregate functions. Through services orchestration model the whole business process is implemented.

All other types of program components are not designed to be interoperable. An interoperable software component should have the following properties:

- it is self-described to be discoverable and usable by other software components
- it runs without interruption from beginning to end, regardless of any inconsistency between the information technologies that are implemented
- it can receive and send messages in a standardized format through which to communicate with other software components
- it does not include human labour that can't be classified as indispensable

In context of opportunities for integration, besides the web services, the most flexible are stored procedures. Modern DBMS provide a wide range of tools for creating stored procedures including embedding different languages and programming technologies and tools to provide existing stored procedures as services. Based on the fact that this procedural element of a database is good enough to describe the ultimate functionality and can easily be made available for use, it can be argued that the stored procedures to integrate most easily.

Factors for determining the appropriate level of interoperability

Before you choose a technology for software components integration, you have to choose which ones and to what extent it makes sense to be restructured as interoperable.

Business factors

- Importance of components from the perspective of the overall activities of the organization. The greater part of the operations automates a software component, the more information it provides to the management of the organization, the more serious investment can be made in the implementation of its interoperability.
- Compliance of the implementation of business logic components. Software components that do not comply sufficiently with the logic of business processes should rather be redesigned and rebuilt than integrated with other components.
- Dynamics of actual business process. The more dynamic a business process is, the more meaningful is its conversion to take the form of one or more services.
- Degree of decentralization of business process - the business processes that are spatially diffused is particularly important to work as interoperable
- Number of current and future users of applications, both human and other

Technological factors

- Openness - it depends on how standardized interfaces that provide component, data structures and messages that can be exchanged, as well as its ability to self-describe.
- Architecture - the more isolated the business logic from the user interface of the software structure is, the higher the degree of interoperability can be achieved.
- Embedded security, as well as options given for authentication and authorization of users of different kinds - people, applications and smart devices.
- Productivity to be achieved in the interaction. It depends significantly on the performance of individual software components taken from the volume of data exchanged and the communication speed

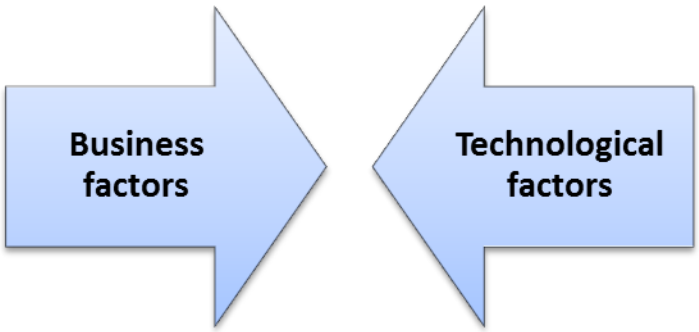


Figure1: Factors for determining the appropriate level of interoperability

An approach to integrate incompatible software components

One option to create an integrated solution to information problems is a complete rewrite of all the "old" (welded) applications using a popular information technology system for building service-oriented (SOA). Apparently, this approach would be extremely costly and lengthy that most companies can't afford.

In recent years, three main technologies for interoperability of processes:

- Infrastructure-oriented messages (MOM – message-oriented middleware). Through this, the programming layer provides asynchronous communication for integrating applications in loosely coupled relationships of the supplier and customer.
- Building Bridges (direct links) between applications. The approach is applied very rarely, only when the performance of the system has a higher priority than interoperability.
- XML web services. XML web services based on standards by which to overcome the barriers posed by different programming environments, operating systems and platforms.

The highest degree of interoperability can be achieved by a selection of software architecture, which consists of XML web services, instead of the mechanisms associated with the remote procedure call. A significant advantage is that the services are self-describing and based on the standards by which to overcome the barriers posed by different programming environments, operating systems and platforms.

With the integration of legacy, applications could apply a combined approach – a selected functionality can be packed as a service by creating a software layer for processing of the messages that activates it and handles the returned result.

A software layer for implementing existing business applications integration

Different types of software components should be used by standard services that make up interoperable business applications of "modern" generation. Their internal organization and functions are different, but their "skin" should look the same, although it is implemented specifically for each type.

One possible approach to the integration of software components of various types into interoperable business application development is to develop a specialized software layer that includes:

- Adapter that first organizes the communication between services and packaged "welded" components, and on the other hand takes care of the overall management of "bundled components" extracting their descriptions, running them according to their extracted description, and receiving the results of the processing and transformation into a standard message structure (e.g. XML)

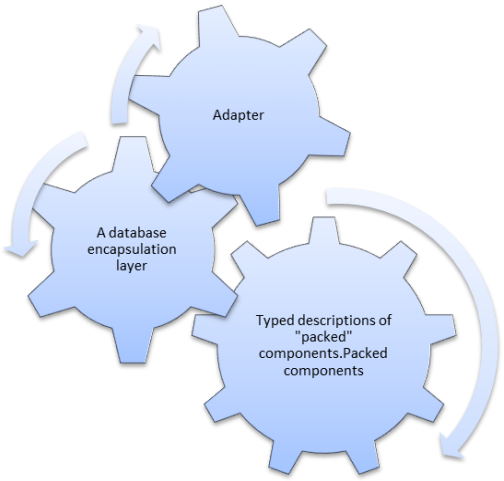


Figure 2: Model of a software layer for implementing existing business applications integration

- Typed descriptions of "packed" components to be integrated. These descriptions should specify name, location, arguments and execution results including possible exceptions of each packed component. Generally these descriptions may be organized like WSDL.
- Packed components including the skins that hide their specificity, so that their invocation during exploitation can be done in a uniform way
- A database encapsulation layer which hides the implementation details of database(s), including their physical schemas, from the business code. This layer provides business objects with persistence services – the ability to read data from, write data to, and delete data from data sources. This layer also provides persistency for descriptions of "packaged" components to be integrated

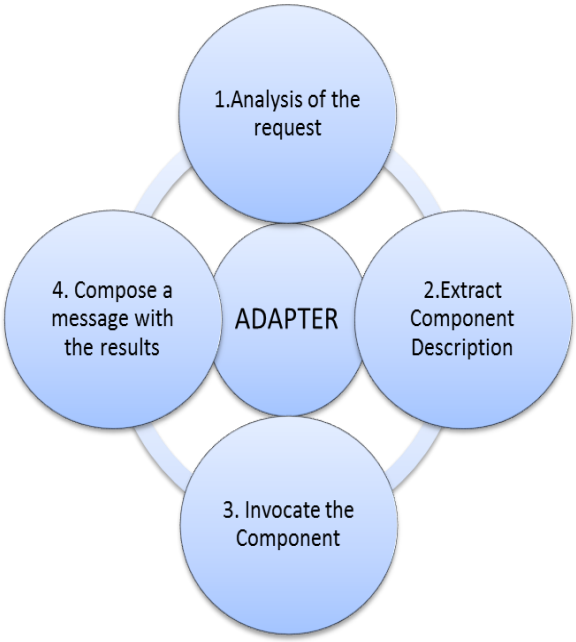


Figure 3: Functionality of the Adapter

Conclusion

Business processes have been and always will be a main reference point for the selection of the architecture of computer systems. The more distributed and interconnected are the business processes, the higher the extent of business applications integration is needed.

The higher extent of interoperability of "existing" business applications is required, the more expensive and complex are process that needs to be undertaken. A highest extent of integration is necessary for software components, with highest importance and usability to the organization. It can be done in cases where there are no unsolvable or unreasonably expensive (outweighing the costs of migration) technological problems.

The next step for implementation of these approaches is to design the "packing" of the various software components that build business applications inherited from the recent and distant past.

References

1. Велев Д., Е. Денчев, К. Стефанова, В. Лазарова, М.Цанева, А. Мурджева. SOA – основно направление в развитието на софтуерните технологии. В: „Бизнес информатика”- сборник доклади от юбилейна международна научна конференция по повод 40-та годишнина на специалност Информатика, С., 11 октомври, 2007 г. стр.78-85
2. Цанева Моника. Избор на степента на интероперабилност на бизнес приложения. „Бизнес информатика”- сборник доклади от юбилейна международна научна конференция по повод 40-та годишнина на специалност Информатика, С., 11 октомври, 2007 г.стр. 166-173
3. Цанева Моника, В. Кисимов, Е. Денчев, А. Мурджева, Д. Велев, К. Стефанова, В. Лазарова. „Интегриране на бизнес приложения – основно предизвикателство към системното програмиране.”, „Бизнес информатика”- сборник доклади от юбилейна международна научна конференция по повод 40-та годишнина на специалност Информатика, С., 11 октомври, 2007 г., стр.215-222
4. An Introduction to Service-Oriented Integration, © 2006 Attachmate Corporation. WHITE PAPER,http://www.csdpcorp.com/LinkClick.aspx?fileticket=mFZGvKbYNLk%3D&tabid=261&mid=965
5. Mark Rascino, The Gartner Scenario: The current state and future direction of IT, Gartner Symposium/ITxpo Africa 2006, Cape town International Convention Centre, Cape Town, South Africa
6. Oracle Agile Product Lifecycle Management, AIS Developer Guide v9.3.0.3 Part No. E23055_01 August 2011 http://docs.oracle.com/cd/E23020_03/otn/pdf/integration/ E23055_01.pdf
7. Scott W. Ambler, Encapsulating Database Access: An Agile "Best" Practice, http://www.agiledata.org/essays/implementationStrategies.html
8. http://webservices.xml.com/

INFORMATION SYSTEM FOR TAGGING, INDEXING AND SEARCH OF UNSTRUCTURED DATA

Maria Marzovanova, Teodora Tsoykova, Valentin Kisimov

Abstract. Piling up large amounts of information in diverse areas of science and business in the form of documents, with no certain rules for their structure, is the main cause of the difficulties in searching and subsequent efficient reuse of that content. Any information that either does not have a pre-defined data model and/or does not fit well into relational tables is referred to as unstructured data. Usually those are customer reports, invention proposals, e-mail messages, draft documents, project plans, operational memos, etc. With the increasing of the role of Web 2.0 technology, where all information is in Web pages, the unstructured data is becoming more and more important. If we add audio, movies and pictures, the amount of the unstructured information is becoming more essential. In the world, more than 70% of the information is in unstructured form. The processing of unstructured data results in irregularities in comparison to the traditional computer programs using stored data in Databases and in structured files. General approach to work with unstructured data is to apply tagging, creating indexes, and special search mechanisms based on the tags. The purpose of the current paper is to propose an Information system based on tagging principles, creating special interfaces to the unstructured data formats, and using dedicated search algorithms for finding the required information.

The paper proposes an ICT architecture of Information system, using unstructured Non-Web and Web 1.0/2.0 data sources, applying Taxonomy of aggregatede tags, creating Indexes according to the Taxonomy, and building the correpondent Search algorithms. The developed Interfaces provide access to the main unstructured data formats – HTML, Word, Pdf, XML, etc., where the access to all those formats is standardized. For aggregation of tags into Taxonomy, a special system is developed, where different relationships between the tags are possible for the creation of an extended taxonomy – parent-child, association, and similarity. The Indexes are a new type of information pointers for the unstructured data, helping the search algorithms. The paper offers a prototype ICT solution, where main functionality is tested.

Keywords: tagging,indexing, search, unstructured data, ICT architecture

Introduction

Piling up large amounts of information in diverse areas of science and business in the form of documents, with no certain rules for their structure, is the main cause of the difficulties in searching and subsequent efficient reuse of that content. Any information that either does not have a pre-defined data model and/or does not fit well into relational tables is referred to as unstructured data. Usually those are customer reports, invention proposals, e-mail messages, draft documents, project plans, operational memos, etc. With the increasing of the role of Web 2.0 technology, where all information is in Web pages, the unstructured data is becoming more and more important. If we add audio, movies and pictures, the

amount of the unstructured information is becoming more essential. In the world, more than 70% of the information is in unstructured form. The processing of unstructured data results in irregularities in comparison to the traditional computer programs using stored data in Databases and in structured files.

Unstructured data and tagging

As most of the definitions for unstructured data explain it in contrast to structured data, it is good to start with clarifying the second term first. Very neat and clear distinction between these two areas of data is given by the authors of the book Text mining [1]. Simply put they write that if you ask a data-mining specialist, as data-mining applications are all about structured data, they see data in numbers. On the other hand if text-mining specialists work with readable content where every word has its obvious meaning.

Structured data in this book is defined as all data that can be transformed by a "data preparation" process or can be collected based on careful prior design for mining. For that purpose it's important that the items that will be used to be clearly described over a range of all possibilities, and these are then recorded uniformly for every example that is a member of the sample. Each attribute has either numerical value that can be compared or it has a code in a numerical format. The meaning of the code is described elsewhere.

Therefore, if data can be described by a spreadsheet with its tabular format, then it is structured. All other data is classified as unstructured. In the area of unstructured data Geoffrey Weglarz finds two major groups – textual object and bitmap objects [2]. People use and produce that kind of data every day and if a proper tool for processing and information retrieval is applied on that data, it would save these people time and increase their efficiency. This paper's proposed system architecture is regarding the needs and specifics of the textual data-division.

The popular approach for organizing textual data is categorization. Two are the best known mechanisms for that – taxonomies and folksonomies. Basically the process in both cases consists of putting some piece of data into a category, that is part of the used structure. Some of the differences are that folksonomies are user-generated, with no control and predefined structure and they are used for categorizing web data. Taxonomies when used for organizing information are usually part of a specific system, they are created by specialist, difficult to customize and with strong rules for modification. They are both usable for what they have been created but in matter of unstructured data management one major disadvantage emerges in all popular systems based on them. The way these systems are developed tagging or classification can't be done without human intervention, people should first read the text and then put it in a category. Moreover it's the same when retrieving the results of a search, people would know if this is the appropriate document only after having spent some time reading it. [3] [4].

Information system architecture

The definition of information retrieval system, is that it is a system that is capable of storage, retrieval, and maintenance of information. Information in this context can be composed of text (including numeric and date data), images, audio, video and other multi-media objects. [5]

Melucci and Baeza-Yates in the book Advanced Topics in Information Retrieval-the full-fledged web search engine would contain three components – a crawler, an indexer, and a query processor [6]

Based on that definitions this report offers an architecture for Information System (**Fig. 1**) containing the following sub-systems – Information system for creating taxonomy/folksonomy and Information system for tagged knowledge search.

The component of the information system for creating and managing taxonomy/folksonomy is responsible for organizing a group of terms in an extended hierarchical structure. The resultant structure is based on hierarchical and relationl taxonomy but aggregating features from folksonomy, thesaurus, ontology and dictionary.

The system for tagged knowledge search is designed to assist users in organizing unstructured content and performing subsequent search queries. In its foundations lie four main objectives:

- Accessing various types of unstructured data sources,
- Tagging specific pieces of content of the unstructured data sources,
- Indexing,
- Search queries execution.

This report is focused on this second component of the bigger system and its architecture. The following sections describe in details how each of the four objectives is fulfilled in the proposed software solution.

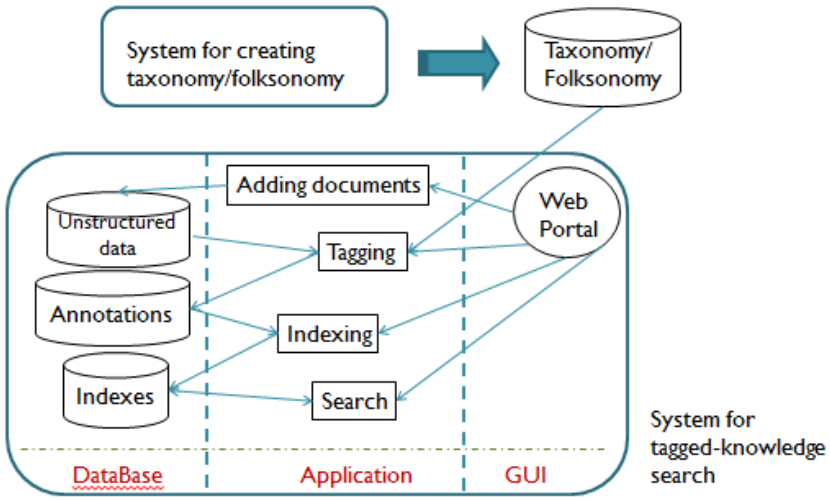


Fig. 1.

Accessing content and tagging

As mentioned above, unstructured data comes in a variety of file formats each having its own structure and representation specifics. The differences between the file formats

influence accessing their content and performing the system procedures on it. This issue leads to the first component of the system – accessing content.

For the system needs its necessary to treat the different unstructured-data sources equally. The solution is to develop mechanism for format recognition and separate interfaces for each of these formats. Once the content is reached it can be extracted to a common format in the system for further use. Such functionality is in the area of text engineering and there are a number of software tools and development environments available that can be of help. Among the popular software dealing with the problem GATE is known to be the most appropriate for the needs of the system. [7]

GATE or General Architecture for Text Engineering [8] is a Java suite of tools originally developed at the University of Sheffield beginning in 1995 and now used worldwide by a wide community of scientists, companies, teachers and students for all sorts of natural language processing tasks, including information extraction in many languages.

The GATE family includes:

- *an IDE* - GATE Developer - an integrated development environment for language processing components bundled with a very widely used Information Extraction system and a comprehensive set of other plugins,
- *a web app* - GATE Teamware - a collaborative annotation environment for factory-style semantic annotation projects built around a workflow engine and a heavily-optimized backend service infrastructure,
- *a framework* - GATE Embedded - an object library optimized for inclusion in diverse applications giving access to all the services used by GATE Developer and more,
- *an architecture* - a high-level organizational picture of how language processing software composition,
- *a process* for the creation of robust and maintainable services.

One of the reasons GATE has lasted well and been successful is that the entire core is broken down into reusable chunks. Using Gate Embedded it is possible to embed language processing functionality in any application.**Fig. 2** summarizes the main APIs available.

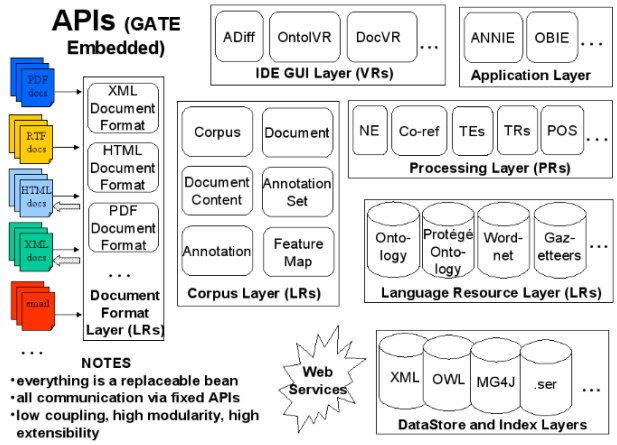


Fig. 2.

Having in mind the outlined objectives of the System for tagged knowledge search the java application embedding gate functionality includes components

Information about all the unstructured-data sources imported for processing is saved in a table in the system's database. The developed Gate application locates these files and unites them to form a GATE Corpus. A Corpus is a language resource in GATE and in fact that is a set of Documents that you can run a process on without caring about the files' actual format.

As they say in the GATE user guide, the basic business of GATE is annotating documents, and all the functionality it offers is related to that. Adding a document in a corpus is basically a preparation for putting some annotations on it. The annotating GATE component used is the ANNIE Gazetteer. Its responsibilities in the system include getting a list of terms, extracted in advance from a certain taxonomy/folksonomy, and annotating every occurrence of them in the content of the corpora documents. As a result of that process an XML file for each document is created. That file has all the content of the original source extended with the tags applied on the terms' occurrences and a list of all added annotations in the end. The list of the annotations is used to extract information about which term in which file is found. That information is then saved in a database table for further use.

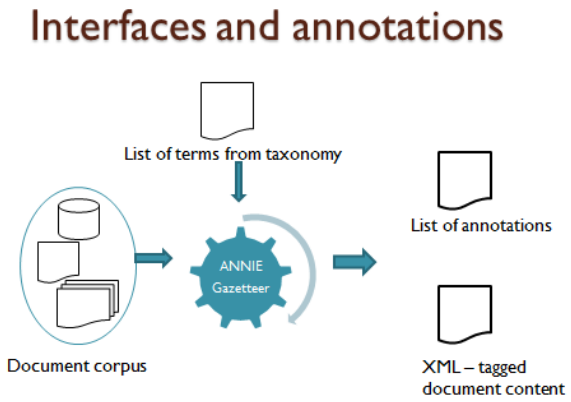


Fig. 3

Indexing and searching

Searching in the system is actually performed over the terms used to build a taxonomy/folksonomy. The results presented in the end are based on the indexes connected to the terms that suit best a certain query entered by the user.

What is called an index in this case is a structure that makes the connection between a term from a taxonomy/folksonomy and its occurrences in the content being searched. It's more like a pointer that consists of two groups of fields holding information about the term and the annotation, respectively (**Fig. 4.**). To identify the term with its specific use in a certain taxonomy it has the term's id and the taxonomy/folksonomy id. The document id and the annotation position constitute the second group of fields.

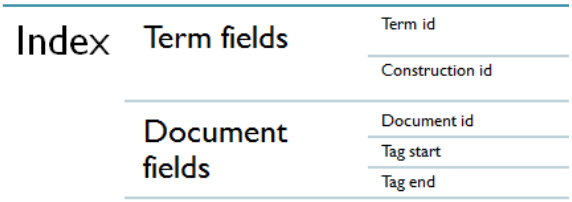


Fig. 4.

The information needed to create an index is taken from the results of the annotating process. The indexing is entirely automated process that starts right after the end of tagging.

The system is designed to provide a wide range of search options. Starting from the mechanisms closest to the database of terms, the first process is analyzing the user's query. The aim at this point is to recognize the terms saved in the system in the entered query. As a result all uses of the entered text or similar to it are visualized, so that the user can choose which one to proceed with. From the system's point of view this gives information about the taxonomy and the exact term use.

Getting a little more distanced from the database of terms, the user can choose between term focused search and subject focused search. The term focused search is limited to results including only the term entered, while subject focused search delivers additional results including terms connected to the entered one. The relationship between the additional terms is defined in the taxonomy and these actually are the terms having a direct relation of any type to the one that the user looked for. In this case documents containing synonyms, translations or details of the needed term will not be omitted.

Conclusion

The paper presents an architectural solution for reengineering of unstructured non-Web and Web 1.0/2.0 data and documents for the purpose of tagging tem, indexing of those data, and providing searching of these data. Special type of taxonomy is used, having hierarchical and relational links, making it as a transition to a ontology. The proposed architecture has been prototyped and tested in UNWE, Bulgaria.

References

[1] S. M. Weiss, N. Indurghya, T. Zhang and F. J. Damerau, "Text mining," Springer.

[2] G. Weglarz, "Two Worlds of Data – Unstructured and Structured".

[3] R. YEE, Pro Web 2.0 Mashups: Remixing Data and Web Services, Apress, 2008.

[4] R. B.-Y. a. B. Ribeiro-Neto, Modern Information Retrieval, Addison Wesley; 1st edition , 1999.

[5] G. J. Kowalski and M. T. Maybury, Information Storage and Retrieval Systems: Theory and Implementation, Springer, 2000.

[6] M. Melucci and R. Baeza-Yates, Advanced Topics inInformation Retrieval, Springer, 2011.

[7] H. Cunningham, Y. Wilks and R. J. Gaizauskas, "GATE - a General Architecture for Text Engineering".

[8] H. Cunningham, D. Maynard, K. Bontcheva and etc, "Developing Language Processing Components with GATE," The University of Sheffield, Department of Computer Science 2001-2012, September 2012. [Online]. Available: http://gate.ac.uk/sale/tao/split.html. [Accessed October 2012].

Architecture of a business intelligent system for monitoring the present and the future state of a database

Alexandrina Murdjeva, Veska Mihova

University of national and world economy
amurdjeva@unwe.bg,vmihova@unwe.bg

Abstract. In today the main problems attendant on the business applications are relevant to their productivity and performance.
The problem of monitoring and productivity optimization of the database of a business application is a traditional problem and for its solution there are many various concepts and software products developed. But all solutions perform monitoring on the current condition of the information system which makes the process of optimization slow and ineffective.
The current report proposes architecture for monitoring the current and the future state of a database and explains the idea of forecast and monitoring the future state of a database.
The proposed architecture would give the business enough time for reaction and timely organization of problem solving.

Keywords: database, database monitoring, database administration, database performance, performance forecast

Introduction

Business software applications' main problems are relevant to their productivity and performance.

Business information systems' productivity is an indicator determining the data processes' execution time and users' response time.

Together with the complex information systems' architecture, a very important indication for the development of the information system is the permanent growth of stored and managed data. This growth worsens the system's productivity and decreases the system's performance.

The solution to the problem with the systems' productivity has already been established, but in some aspects there are serious weaknesses provoking the research better solutions.

The problem of business information systems' monitoring and optimization has been a current topic in recently because the need for a solution is recognized in more and more business areas.

State of the art

The problem

With the development of new technologies in the modern world, the work of a business organization is monitored by the so-called monitoring systems. The monitoring of an

information system's operation and database monitoring in particular is a task which has been defined and implemented for a long time. The problem of monitoring and productivity optimization of the database of a business application is a traditional problem and for its solution there are many various concepts and software products developed.

The basic concepts are developed in two main directions: monitoring of the condition and providing opportunities for change. Most of the modern database management systems (DBMS) integrate various implementations of these concepts in the form of software components, system elements and integrates external ones.

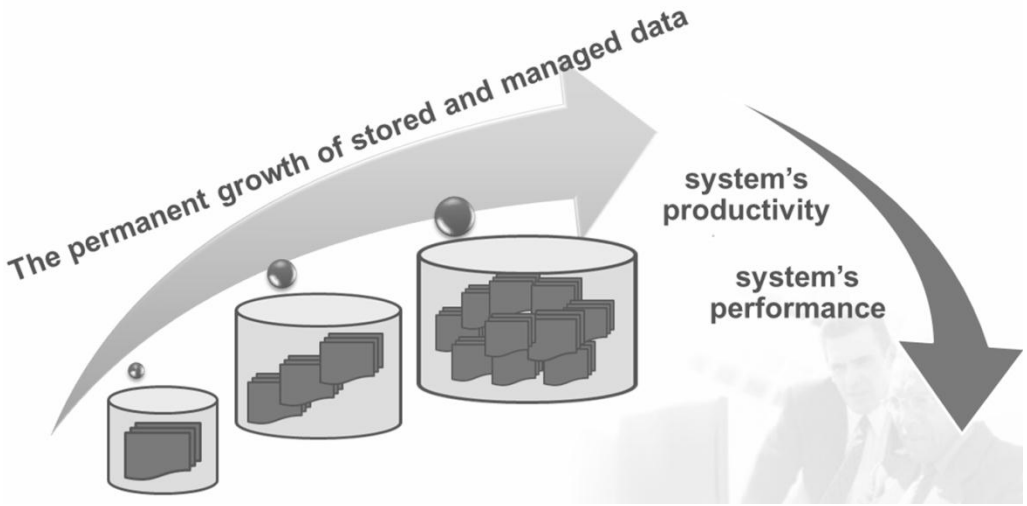


Fig. 1. The problem

Literature review – monitoring systems

The monitoring systems propose consolidated information that displays the present status of an information system.

These reports notify the database administrators and system administrators for potential problems. Thereby the reports help administrators to manage the system better.

The common technique among the existing solutions for performance management problems is system monitoring. Most of database management systems' manufacturers have implemented functionality for the activity monitoring in their software products.

The common functionalities and characteristics that monitoring systems propose can be summarized:

- Monitoring systems propose tracing system's activity. They trace a set of execution characteristics and health indicators of the system. Indicators like database calls, execution time of the queries, number of rows, active sessions, system resources – occupied RAM, CPU status and many others.
- Monitoring systems propose creating and editing user defined indicators for monitoring.
- Monitoring systems propose editing and inserting system's bottlenecks.
- Monitoring systems propose generating a signal for occurred problem with the system.

- Monitoring systems propose receiving a signal in real time when some of the indicators exceed some bottleneck.
- Monitoring systems propose tools for visualizing of the monitoring values.
- Monitoring systems propose consolidated summaries which show the present status of the system.
- Some monitoring systems propose analyzing of some queries and transactions and eventually solutions for improved performance in the present moment.

The summarized characteristics are found in known tools as for example:Oracle Enterprise Manager (5),

- Applications Manager - Database Query Monitor, (6)
- SQLDBManager Plus, (7)
- DB Monitor Expert (8)
- AppSQL Xpert, (9)
- Qure, (10)
- Други.

After the performed literature review of the existing systems, a search for better solutions that give administrators enough time for reaction is inevitable.

Fig. 2. The present state of the problem

Results

The idea covers tasks and concepts that are very similar to those of business intelligent systems - data collection, storage history, producing new values based on the accumulated history, forecast. That's why, this report takes the idea of the architecture of business intelligent systems and develops it further to serve the idea of future forecasted environment, monitoring in forecasted state of the database.

This idea can be realized with the help of a business intelligent system which would have the following architecture:

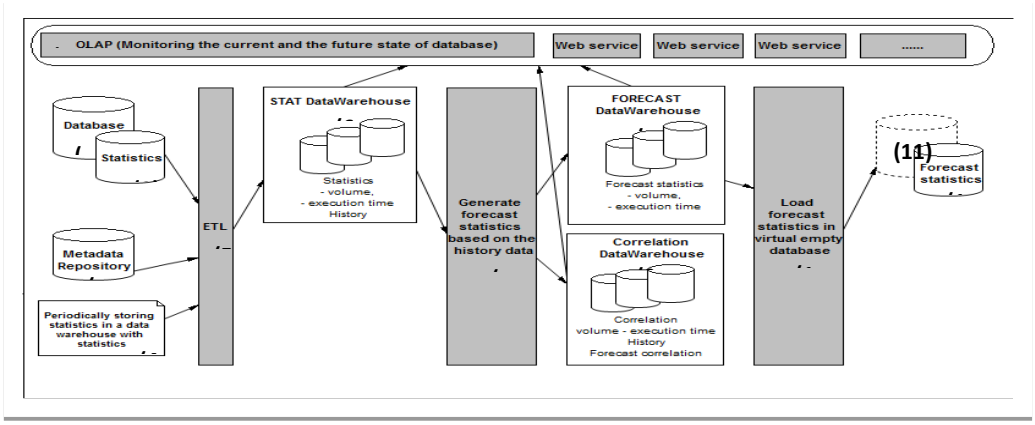


Fig. 1. General BI architecture for monitoring the current and the future state of a database

Fig.3. Architecture of a business intelligent system for monitoring the present and the future state of a database

The data source in the proposed architecture consists of the database of the business information system (1) and the database that stores metadata (3) for the management process of generating the structure of the data warehouse. This metadata is stored with a set purpose - automatic generation of data warehouse according to the specifics of the database. The idea is to generate a data warehouse with a structure that is appropriate to a different DBMS. The data warehouse is designed to store historical information that is extracted from the system tables and the tables of statistics of a particular DBMS.

The data warehouse is loaded by ETL processes (4, 5). The data warehouse (6) is the structure of tables that stores statistics for the execution time of queries, data volumes in different database objects and other necessary relevant statistics for generating forecast values.

ETL tools, stored procedures, offline processing or other appropriate tools (4, 5) provide generation capability of statistical values at a precise time, depending on how dynamic the process of system operation is.

After collecting enough historical information about the operation of the system (6), predicted values (8) are generated periodically (7) based on the gathered statistic.

The generation (7) of these forecasted statistics can be done using statistical packages, Data Mining, stored procedures, MS Excel and many other instruments.

In the proposed architecture a data warehouse component with forecasted data is provided. This is still a storage structure (8), since this is historical data, accumulated values for different periods, with the only specific that these periods are in the future. Apart from data warehouse with forecasted information about data values that are stored in the system and the execution time of queries, the architecture makes the collection of correlation statistics (9) between data volumes and the execution time of queries for a certain period of time possible.

The correlation coefficient between the two variables is calculated and stored within the data warehouse (9).

As the survey shows, at different moments of time one query is affected by different resources. The idea is that BIS can accumulate historical information about these dependencies and, within a particular time, can record their correlation value.

After enough forecasted database statistics are generated using stored procedures, for example, these forecasted values (12) will load (10) in a new empty virtual database (11). This means that actually there will be no data in the database. It will be loaded only with forecasted database statistics.

The focus is on having no data in the database (11). The idea is to provoke the query optimizer to propose an execution plan for the monitored queries according to the estimated future state of the database.

Considering that monitoring systems are based on statistical data, monitoring of the future state of a database (13) can be performed. Of course, the monitoring can be realized using different instruments or systems, for example using OLAP tools.

Monitoring is a process that may occur irregularly and remotely, which requires consideration of technological capabilities to create a suitable user interface for presenting the available information. This interface should be appropriate for isolating the specifics of the overall architecture to enable a choice of different technical realizations, depending on the specifics, needs and policies of the organization. Creating a layer of services to be used as interface for accessing data in the overall architecture is a good approach.

Furthermore, the architecture offers an additional layer of web services (14) which provides the necessary data for monitoring the future state of the database so that any business organization can integrate, visualize and implement the data in a different and appropriate way.

Conclusion

In summary, in this paper an architecture for monitoring the future state of a database is presented. In the beginning the authors present the problems with the business applications. Then a literature review for monitoring and database monitoring is made. The authors propose architecture of business intelligent system for monitoring the future state of a database. The proposed solution gives the business enough time for reaction to an occurring problem with the application's performance and respectively the database's performance.

Fig. 4. Better database's performance

As for future work, the authors intends to develop and present a tool using the concept of this paper.

References

1. [Онлайн] http://www.manageengine.com/products/applications_manager/databa se-query-monitoring.html.

2. [Онлайн] http://docs.oracle.com/cd/B10500_01/server.920/a96533/statspac.htm.

3. [Онлайн] http://en.wikipedia.org/wiki/Database_activity_monitoring.

4. [Онлайн] http://www.oracle.com/us/products/applications/ebusiness/overview/ index.html.

5. Oracle9i Database Performance Tuning Guide and Reference Release 2. [Онлайн] http://docs.oracle.com/cd/B10501_01/server.920/a96533/stats.htm.

6. Product Feature Comparison: ManageEngine SQLDBManager Plus vs. Applications Manager. [Онлайн] http://www.manageengine.com/products/ applications_manager/sql/compare-database-tools.html.

7. Lane, Adrian. Database activity monitoring keeps watch over your data. *http://searchsecurity.techtarget.com.* [Онлайн] http://searchsecurity.techtarget.com/ article/Database-activity-monitoring-keeps-watch-over-your-data.

8. Database monitoring best practices: Using DAM tools. [Онлайн] http://searchsecurity.techtarget.com/tip/Database-monitoring-best-practices-Using-DAM-tools.

9. Proctor, Paul E., Litan, Avivah и Nicolett, Mark. Pattern Discovery With Security Monitoring and Fraud Detection Technologies. [Онлайн] 2 September 2009 r. http://www.gartner.com/DisplayDocument?id=1160514.

10. Quarles van Ufford, Deborah. Business Intelligence: The Umbrella Term, VU Amsterdam. [Онлайн] 2002 r. http://www.few.vu.nl/stagebureau/.

11. Yuhanna, Noel, Penn, Jonathan и Smillie, Katie . Forrester Research:Research: The Forrester Wave™: Enterprise Database Auditing And Real-Time Protection, Q4 2007. [Онлайн] October 2007 r. http://www.forrester.com/
The+Forrester+Wave+Enterprise+Database+Auditing+And+RealTime+Protection+Q4+2007/fulltext/-/E-RES41970?objectid=RES41970.

12. Михова, Веска и Мурджева, Александрина. *Бизнес интелигентни системи за оптимизиране и управление на приложни системи.* София : н.а., 2010. Международна конференция "Комуникации, информационни технологии и статистика. актуални проблеми на теорията и практиката".

13. Мурджева, Александрина и Михова, Веска. *Мониторинг и анализ на бъдеща производителност на бизнес приложения.* София : ЮБИЛЕЙНА НАУЧНА КОНФЕРЕНЦИЯ„СТАТИСТИКА, ИНФОРМАЦИОННИ ТЕХНО-ЛОГИИ И КОМУНИКАЦИИ” 24-25, 11„ 2011.

Metadata for generating a specific data warehouse

Veska Mihova, Alexandrina Murdjeva

Bulgaria, Sofia, University of national and world economy
vmihova@unwe.bg, amurdjeva@unwe.bg

Abstract. In today the main problems attendant on the business applications are relevant to their productivity and performance.
The problem of monitoring and productivity optimization of the database of a business application is a traditional problem and for its solution there are many various concepts and software products developed. But all solutions perform monitoring on the current condition of the information system which makes the process of optimization slow and ineffective.
The current report proposes a metadata model for manage generating a specific data warehouse and explains the idea of forecast and monitoring the future state of a database.
The proposed model would give the opportunity for isolating the specifics of different databases and also would give an easy way to organize the process of generating the specific data warehouse.

Keywords: metadata, database monitoring, database administration, database performance, performance forecast

Introduction

Business software applications' main problems are relevant to their productivity and performance.

The problem of business information systems' monitoring and optimization has been a current topic recently because the need for a solution is recognized in more and more business areas.

The seeking for solution in this area – improvement of systems' performance, is still in progress.

This paper is focused on a little part of a concept for performance management – generating a specific data warehouse.

State of the art

Problem

If we consider the information system as a whole entity, composed of great number of elements, there can be found performance problems on every layer in the architecture of the system.

- Database layer,
- Business logic layer
- Presentation layer

Every layer in an information system has its specifics and its weaknesses. The system layers are interrelated. If a problem occurs in one of the layers, the problem will become a problem for the whole system.

Literature

Monitoring the performance of information systems is related to collecting data for occurring events (applications, processes, etc.). Modern technologies allow monitoring the performance to be made at each layer of the architecture of the system but the highest level of data provision has the layer for data management.

This data provision is better due to the database management systems' supply of large number of data and statistics for the functioning of the system.

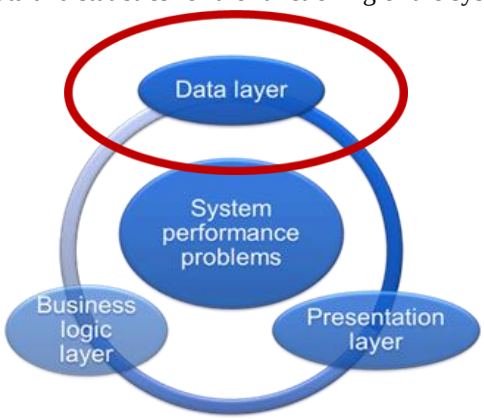


Fig. 1. Where can in an information system occur performance problems?

The information about the performance of a system may be provided from the database management system itself. Each system provides system tables and views which contain information about the structure of the database, its work and the work of the system.

This information may be used for monitoring, as it is available at different levels of detail – individual values, cumulative values.

For performing qualitative monitoring, the information that should be collected is for:

- Data volume;
- The execution of queries;
- The execution time of queries;
- Execution plans;
- Database users;
- Loading of the system, i.e. simultaneous operation of multiple users, number of user connections to databases at different times;
- Time to connect to the database;
- Number of queries to the database;
- Regularity and frequency of executions of queries;

- The database structure (tables, columns, relations, etc.);
- Number of created indexes;
- The operation of the cache buffer;
- The execution of transactions;
- The number and duration of execution of current operations;
- The number and duration of sessions;
- Number and duration of the current lock resources;
- And others.

This information is available in each database management system and it represents the current status of database including the collected statistics which are a picture of the database. The larger DBMS (database management system) provide functionality for collecting historical information, but with significant restrictions. Therefore the layer of data management in the whole architecture of the business application is well provided with information. This is a good opportunity for collecting historical information, which could be used for forecasting the future state.

Results

In order to perform data forecasting, there have to be enough historical data gathered as a basis.

As we mentioned above, every DBMS provides information about the work of a database, but these values are a moment picture of the database state and are not being stored. This arises the need for proposing a concept for monitoring the future state of a database, as well as the need for a well structured data warehouse being the foundation of this concept.

The data source in such architecture for monitoring the future state of a database consists of the database of the business information system and the database that stores metadata for the management process of generating the structure of the data warehouse. This metadata is stored with a set purpose for automatic generation of data warehouse according to the specifics of the database. As the process of generating structures for a data warehouse can be automated, we propose storing metadata to help controlling it. It follows a model of metadata for managing the generating process and the generation of specific data warehouse:

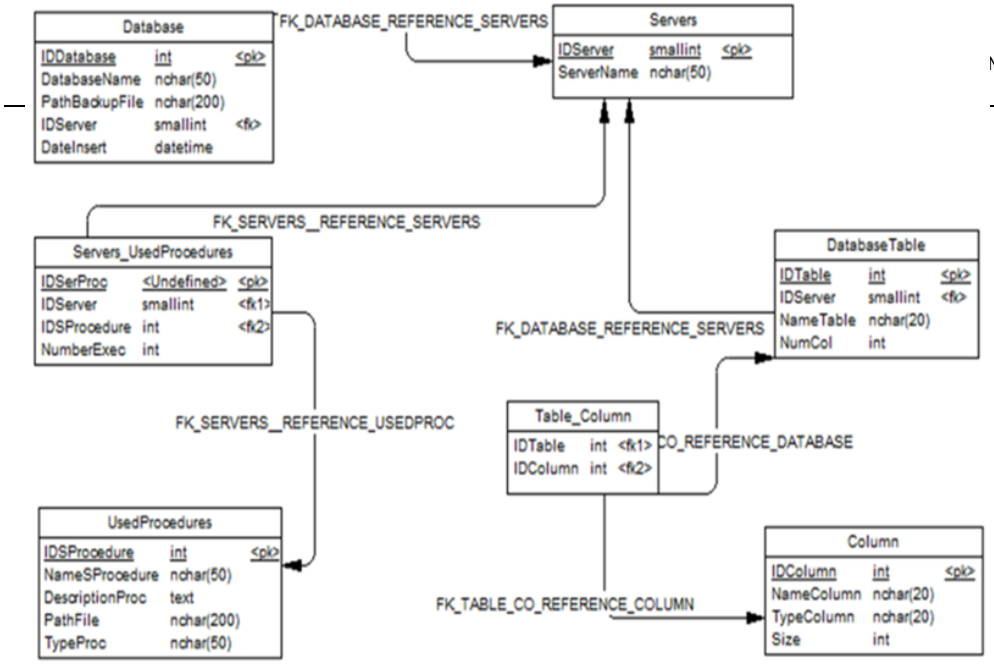


Fig. 2. Metadata model for generation management of a specific Data Warehouse

This metadata can indicate what the corresponding database's name is, what DBMS it works with, and the path to the database, and other relevant data. There are a set of stored procedures that are used for managing the whole warehouse generating process. The idea is to generate a data warehouse with a structure that is appropriate to a different DBMS.

The different types of DBMS store information about the work of a particular database in diverse in name and structure tables and views. This arises the need for storing metadata for each type of DBMS.

The data warehouse is designed to store historical information that is extracted from the system tables and tables of statistics of a particular DBMS so its data source is potentially one and is homogeneous. Therefore it's reasonable to propose storage with an analogous structure to the system tables, equipped with an indicator of time. In such a structure it is possible to provide procedures for generating these structure-based meta descriptions.

Conclusion

In summary, in this paper metadata model for generation management of a specific data warehouse is presented. In the beginning the authors present the performance problems on each layer in business applications. Then a literature review for data provision is made. The authors propose architecture of business intelligent system for monitoring the future state of a database that requires generating a specific data warehouse. The proposed solution gives business a specific data warehouse for storing statistical history data, which helps to manage the application's performance and respectively the database's performance.

Fig. 3. Better database's performance

As for future work, the author intends to develop and present a tool using the concept of this paper.

References

1. Михова, Веска и Мурджева, Александрина. *Бизнес интелигентни системи за оптимизиране и управление на приложни системи.* София : н.а., 2010. Международна конференция "Комуникации, информационни технологии и статистика. актуални проблеми на теорията и практиката".

2. Product Feature Comparison: ManageEngine SQLDBManager Plus vs. Applications Manager. [Онлайн] http://www.manageengine.com/products /applications_manager/sql/compare-database-tools.html.

3. [Онлайн] http://www.manageengine.com/products/applications_manager/ data base-query-monitoring.html.

4. [Онлайн] http://www.oracle.com/us/products/applications/peoplesoft-enterprise/ overview/index.html.

5. Brackett, M. Business Intelligence Value Chain. [Онлайн] 1999 г. http://www .dmreview.com/editorial/.

6. Eckerson, W., Howson, C. Enterprise Business Intelligence: Strategies and Technologies for Deploying BI on an Enterprise Scale (Report Excerpt). [Онлайн] 2005 г. http://www.tdwi.org/Publications/WhatWorks/.

7. Lane, Adrian. Database activity monitoring keeps watch over your data. *http://searchsecurity.techtarget.com.* [Онлайн] http://searchsecurity.techtarget.com/ article/Database-activity-monitoring-keeps-watch-over-your-data.

8. Database monitoring best practices: Using DAM tools. [Онлайн] http://searchsecurity.techtarget.com/tip/Database-monitoring-best-practices-Using-DAM-tools.

9. Мурджева, Александрина и Михова, Веска. *Мониторинг и анализ на бъдеща производителност на бизнес приложения.* София : ЮБИЛЕЙНА НАУЧНА КОНФЕРЕНЦИЯ„СТАТИСТИКА, ИНФОРМАЦИОННИ ТЕХНОЛОГИИ И КОМУНИКАЦИИ" 24-25, 11,, 2011.

10. Proctor, Paul E., Litan, Avivah и Nicolett, Mark. Pattern Discovery With Security Monitoring and Fraud Detection Technologies. [Онлайн] 2 September 2009 г. http://www.gartner.com/DisplayDocument?id=1160514.

11. Oracle9i Database Performance Tuning Guide and Reference Release 2. [Онлайн] http://docs.oracle.com/cd/B10501_01/server.920/a96533/stats.htm.

12. [Онлайн] http://docs.oracle.com/cd/B10500_01/server.920/a96533/ statspac .htm.

13. [Онлайн] http://en.wikipedia.org/wiki/Database_activity_monitoring.

STRUCTURED INTEGRATION APPROACH BASED ON LEVELS OF REUSABILITY

S. KOUZMANOV, MONIKA TSANEVA

Department of Information Technologies and Communications, Faculty of Applied Informatics and Statistics, University of World and National Economy

Abstract. Recently the approaches of integrating software systems in heterogeneous environments undergo significant changes with the adoption and development of service-oriented technologies. The service oriented software development offers considerable advantages in the field of software integration, since it is based on open and widely applied standards for interoperability, data exchange in common formats and open data transport protocols. Although service oriented architecture is applied in many projects with different level of interoperability delivery and shows very high rate of success, there is a lack of structured approach which can be applied in wide range of integration problems. This paper proposes adopting a structured approach for executing service oriented integration, which can optimize the cost of implementing such projects and can guarantee a certain level of success.

Keywords: Integration, service-oriented architecture, integration methodology

Introduction

Since service-oriented technologies are based on open standards and can be executed on various platforms, they are successfully used in wide range of applied projects to shorten the gap when software systems built on different proprietary standards should be integrated. Such projects are integration of legacy systems with newly developed software, building a common enterprise integration environment and so on. Development of integration environments based on service-oriented technologies if often technologically successful, but a structured methodology for running such kind of project is not sill adopted. This leads to trying different ways for applying service oriented integration in different projects which doesn't guarantee optimal development time and cost and sometimes the result may not deliver the targeted level of interoperability. For such structured method to be developed the integration environments can be divided in layers and common reusable patterns which are used in applied projects can be identified on each layer.

Definition of Integration Environment

Integration of software systems was a common but a challenging job in software engineering for some time now. An integration project is always related with tasks such as analysing the already existing systems which should be integrated, researching which is the best technology to deliver systems interconnection and functional interface exposure, implementing the system integration interfaces and middleware and, most importantly, defining and delivering the capabilities that should be provided by the integrated solution as a whole.

Since the existing systems which should be included in the integration process can vary significantly from one to another and every system has its own technology and semantic specifics, the environment can become very heterogeneous, both from a technical and abstract perspective. This leads to a more complex implementation phase, which is related with building a dedicated integration interface or the so called adapter for every application in the solution and implementing a common infrastructure, also called middleware, which is responsible for aggregating system functionality and data semantics and last but not least making all of these components to work together.

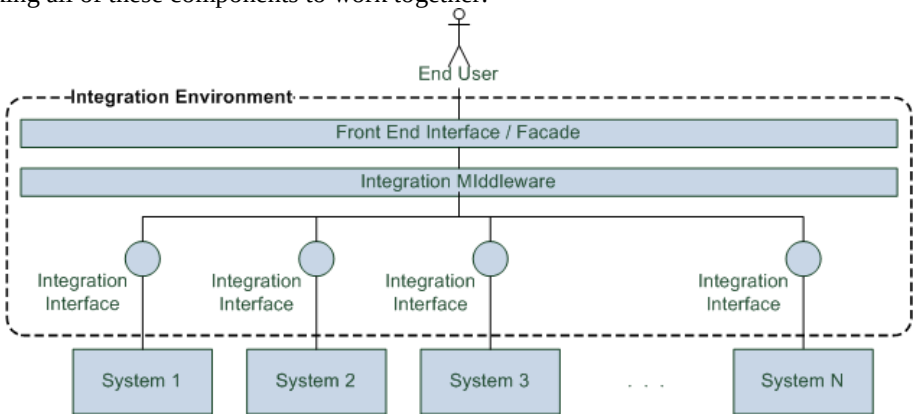


Figure 1. *Integration Environment.*

On that foundation we can define an integration environment as all components between the single systems participating in the integration and the end-user, which includes the system adapters, integration middleware and the interconnection between these modules.

An integration environment can be also defined with the following features:

- It provides software adapters and builds integration interfaces for a common functional exposure.
- It provides a common middleware which aggregates system functionality and data semantics from the stand alone systems.
- It has well-defined composite capabilities and provides a facade or front end interface for consuming the aggregated functions as a single composite application, i.e. the environment provides a certain level of transparency.

Layers of an Integration Environment

Since the aim of our research is to define integration methodology based reusability levels we should firstly divide the integration environment on layers. Such layers are the wrapping (adapter) layer, the data integration layer, the data transport layer, the functional interfaces layer, etc. Each layer has its own techniques for delivering interoperability and they bring different rate of reusability depending on the layer. In that case each layer can be related to a different level of reusability.

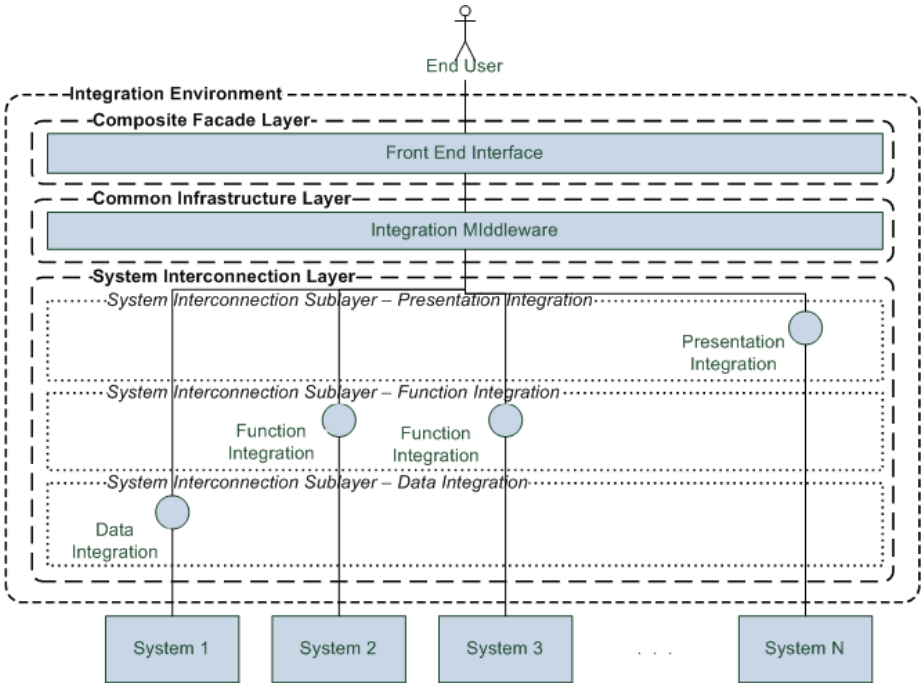


Figure 2. *Layers in an Integration Environment.*

Based on our previous definition of an integration environment we can define the following layers (listed from bottom to top in the stack sequence):

- *System Interconnection Layer* – this layer provides the integration interfaces between the core of the integration environment and the stand-alone systems. It is sometimes called adapter layer, since the system entities are wrapped by the so called adapters – software module which enable their communication and data exchange with the integration solution. With the adoption of service-oriented technologies such adapters are often implemented as software service modules.
- *Common Infrastructure Layer* – this layer provides the integration infrastructure or, in other words, the core services that should be provided to the stand-alone systems. Such services are the functionality aggregation, data transformations, providing common data formats, and communication protocols and so on.
- *Composite Facade Layer* – this layer can be mapped to the presentation layer in a classical software architecture. The difference is that here the front end is not limited to a user interface only. Since an integration solution can be consumed from both users and machines, this facade can be composed from a human interface (for example a portal solution) or a machine consumable interface (for example an orchestrated service set, composite application/service, a business process execution environment, etc.)

As it can be seen from the above layers' definition, the most heterogeneous one is the bottom layer. Since this layer consists of different software modules (adapters), and every

system has its own specifics which should be considered in the wrapping process, it is composite in his nature. To simplify the research of this layer and easily identify some reusable patterns, we can further divide in to sub layers:

- *Data Integration Sublayer* – this layer consists from the adapters used for data integration with stand-alone systems.
- *Functional Integration Sublayer* – this layer consists from the adapters used for functional integration with the system entities.
- *Presentation Integration Sublayer* – this layer consists from the adapters used for presentation integration with lower level systems.

Reusable Patterns

The next step in our research is to identify some reusable implementation patterns which are widely adopted in applied projects. Based on our previous definition of layers in an integration environment we can depict some patterns that can be observed in such kind of solution, starting from bottom to top in the layer stack.

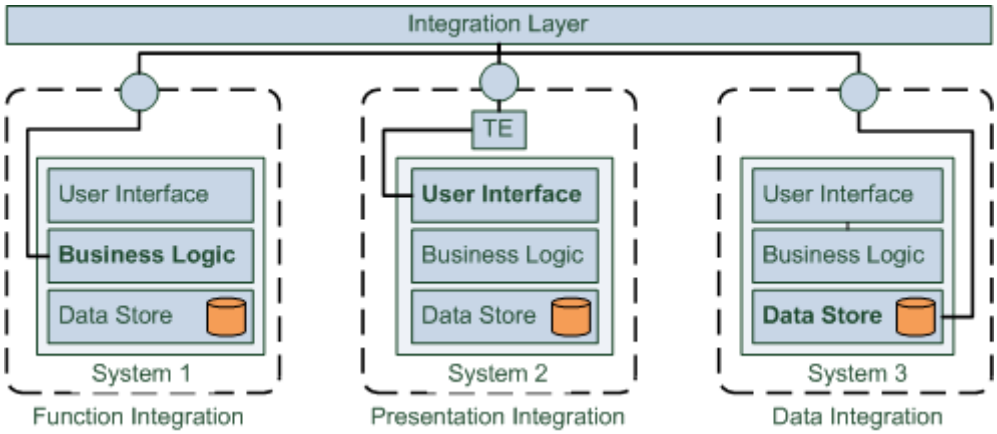


Figure 3. *System Interconnection Patterns.*

First of all, the system interconnection patterns can be matched to the interconnection sublayers and therefore can be easily defined:

- *Data Integration Pattern* – the adapter is implemented to wrap directly the data repository or data-access layer of system that should be interconnected to the integration environment.
- *Functional Integration Pattern* – the adapter is implemented to wrap the business logic layer of the stand-alone system. This can be done by invoking an existing API or if access to the system implementation is available.
- *Presentation Integration Pattern* – if no other options are available the adapter is implemented to emulate an interaction with the user interface of the stand-alone system, or the so called "screen scraping". This pattern has the least invasive implementation.

On the common infrastructure layer we can identify some patterns related with the aggregation data semantics and the dispatching capabilities of this module.

The *Entity Aggregation Machine* pattern can be observed when the integration middleware is used to solve semantic conflict between data fetched from different repository. A semantic conflict appears when data with common semantics has different persistence model across the different repositories. To solve this, the integration middle tier aggregates the fetched entities and outputs the data in a common model.

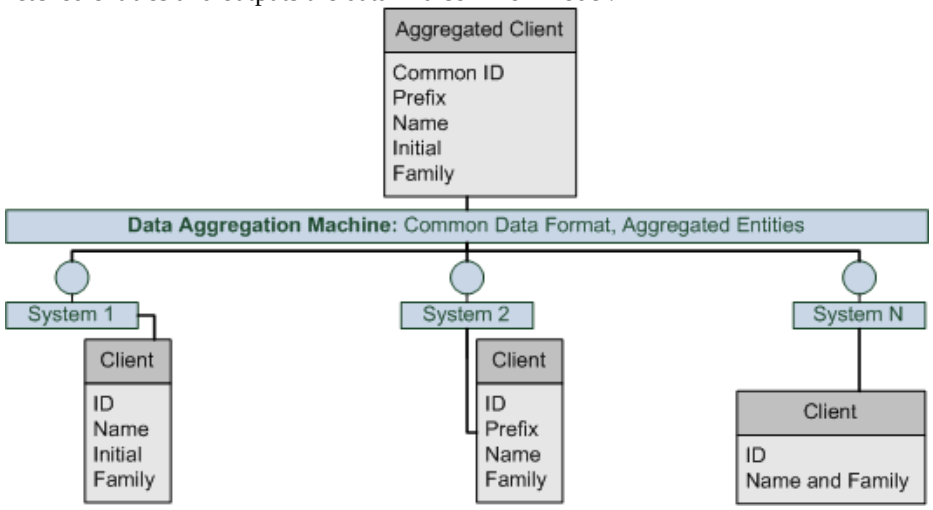


Figure 4. *Data Aggregation Pattern.*

A *Central Broker* pattern can be observed when a central mediation element is implemented to reduce cohesion between the integrated modules/systems. The broker should provide capabilities for routing messages between modules, supporting a systems endpoints registry and data transformation between systems.

A *Service Bus* pattern also has its aim to reduce the coupling between the integrated stand-alone systems. A service bus usually provides a common data-transmission format, bus-command message set and shared infrastructure and protocols for communication. This pattern is usually implemented with web service oriented technologies, which almost naturally provide the capabilities listed above.

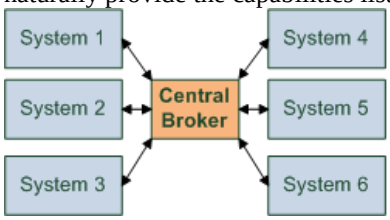


Figure 5. *Central Broker Pattern.*

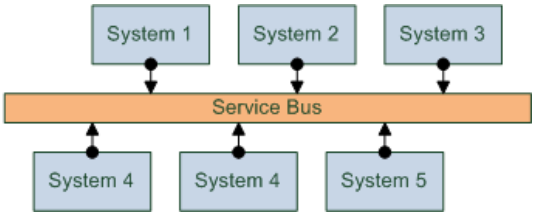


Figure 6. *Service Bus Pattern.*

On the composite facade layer we can identify some patterns related with the presentation of the aggregate results and outputs of the integration solution as a whole.

The *Portal Machine* pattern is implemented when a user needs to interact with more than one stand-alone system to execute his tasks. To avoid manual switching and transition of data between systems, they can be wrapped as services to aggregate their functionality in the integration middleware. If this implementation is available, a common user interface can be presented for a transparent interaction with all the systems.

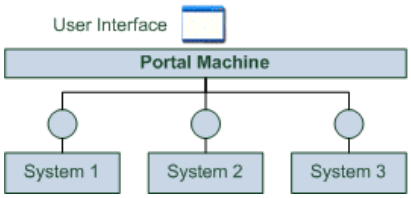


Figure 7. *Portal Machine Pattern.*

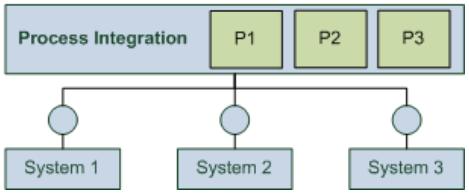


Figure 8. *Process Integration Pattern.*

The *Process Integration* pattern is observed when there is a need for process management. When the single machine tasks in a business process are already executed by existing systems, they can be also wrapped as services and then orchestrated in a process by a process execution middleware (for example a business process management software). Then a user or machine interface for process execution can be provided as a common output of the integrated solution.

Rate of Reusability

The main aim of defining and using patterns in software design and development is to build maintainable software and get predictable results thanks to the pattern reusability and its proven behaviour. After we have defined the layers in an integration environment and the patterns which can be used on each layer we should further investigate their reusability, in order to map each layer to a level of reusability.

To accomplish this we can define *reusability rate* as a measurable property of a pattern. The reusability rate for each pattern can be measured as a sum of key application indicators. Here are some proposals for such indicators:

- *Application* – this indicator can be measured by empirical research to examine in how many already executed integration projects the pattern is applied. For example the function integration pattern can be expected to have higher rate of application than presentation integration, since the first one can be used for more complex tasks.
- *Implementation Effort* – the effort that should be spent by integration developers to implement the pattern. Here also can be applied experimental to approach to evaluate each pattern, but obviously patterns included in "off the shelf" software (for example the service bus solutions) will have better rate on this indicator.
- *Invasiveness* – the more core function of a stand-alone system we access to implement a pattern, the more invasive is the implementation. For example a screen scrapping adapter is the least invasive one, while a function adapter or data adapter is more invasive since it access core functions.
- *Cohesion/coupling* – if the pattern provides lower level of coupling between the elements it will be more reusable, since it can be easily applied with different

modules and within heterogeneous environments. For example the service bus and central broker patterns has lower level of cohesion than a mesh topology.

- *Platform Dependency* – if the pattern leverages more platform independent technologies (like the web-services standards), the more applicable will it be in different environments.
- *Number of off the shelf solutions* – if the pattern is applied in more off the shelf integration middleware packages it will has higher rate of reusability.

Based on these application indicators we can define each pattern's rate of reusability and then we can calculate the average level of reusability of each of the defined integration environment layers. A common integration methodology can be based on these levels of reusability, with the patterns and layers with higher reusability rate will have broader impact on such approach.

Conclusion

After we have defined an integration environment and its common layer that are naturally used in applied projects and after we have identified the most applicable patterns, we have proposed a methodology for evaluating the reusability level of each pattern and layer. Based on these results we can further research the topic to provide calculations for the reusability rates of each component and use them as a foundation for a structured integration approach in which the higher reusability rates will have broader impact on the methodology as a whole.

References

1. Tsaneva Monika. *Selection of Business Applications Interoperability Level*, Business Informatics - Proceedings of the jubilee international academic conference for the 40th anniversary of the Informatics specialty, Sofia, 11 October, 2007, page 166-173

2. Velev D., E. Denchev, K. Stefanova, V. Lazarova, M. Tsaneva, A. Murdjeva. *SOA – Main Field in Software Technologies Development*, Business Informatics - Proceedings of the jubilee international academic conference for the 40th anniversary of the Informatics specialty, Sofia, 11 October 2007, page 78-85

3. Tsaneva Monika, V. Kissimov, E. Denchev, A. Murdjeva, D. Velev, K. Stefanova, V. Lazarova. *Integration of Business Applications – main challenge in system programming*, Business Informatics - Proceedings of the jubilee international academic conference for the 40th anniversary of the Informatics specialty, Sofia, 11 October 2007, page 215-222

Information aspects of local administrative policy

Katya Strahilova

University of national and world economy
katia_emilova@yahoo.com

Abstract. Developing and implementing local administrative policies is a major issue for regional development. In this context, the report examines the issues of application of information technologies as part of local administration and policy. This is important for the extension of the electronic services for citizens and businesses, as well as for the development of the concept of information society. On the basis of analyzes made in the paper, conclusions and directions for future development are formulated.

Keywords: Information technologies, local administration, e-administration.

Introduction

Financial stability is a major problem for local government. Recent years we have seen the expansion of the procedures of decentralization and delegation of authority. [1] These trends lead to increased revenue and spending powers of the municipalities. This necessitates the application of techniques for balancing administrative and professional staff of the municipal administration on the one hand and the new powers on the other hand. Also, we must pay attention to the changes in technical and technological means for taking a new income and expenditure authority derived from the decentralization process

Information aspects of local administrative policy

The policy of providing administrative services at the local level is constantly changing. In this paper, we will introduce those who are related to (Fig. 1):

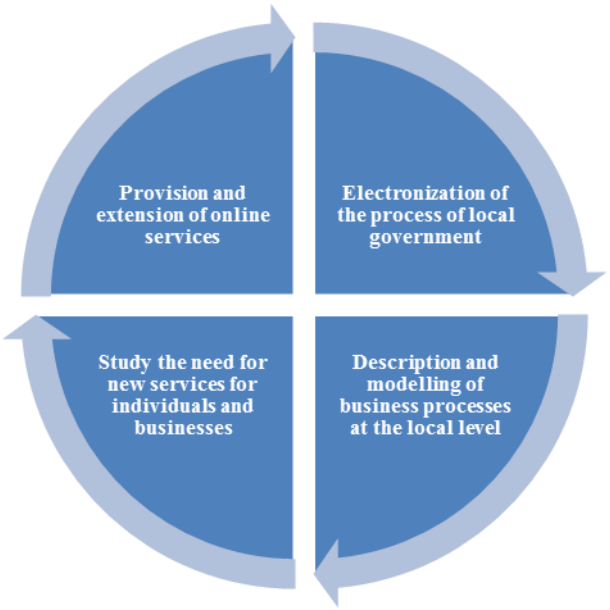


Fig. 1. Information aspects

- Electronization of the process of local government;
- Description and modelling of business processes at the local level;
- Study the need for new services for individuals and businesses;
- Provision and extension of online services, etc.

In this study, under the term "electronic local government" we understand the created and used set of specialists in municipalities, hardware resources and relevant software systems that makes it possible for the municipality to provide electronic services to citizens and businesses. Examinations that we made give us the reason for proposing the following sequence of steps for electronization of the process of local government (Fig. 2):

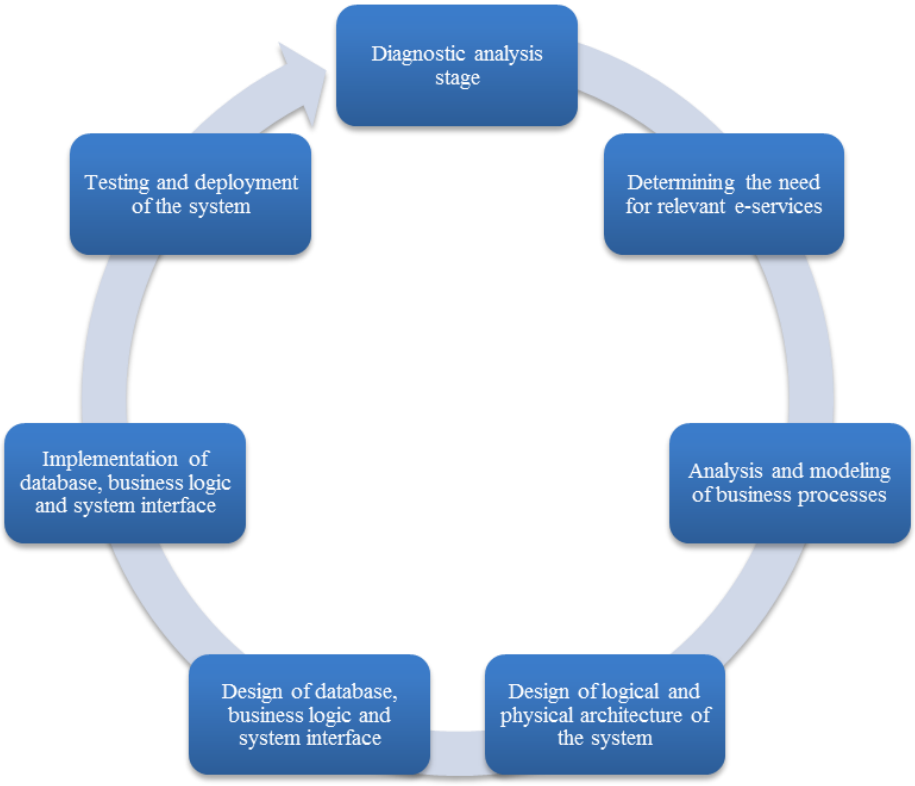


Fig. 2. Electronization of local government

- Determining the need for relevant e-services;
- Analysis and modelling of business processes;
- Design of logical and physical architecture of the system;
- Design of database, business logic and system interface;
- Implementation of database, business logic and system interface;
- Testing and deployment of the system.

As can be seen from the mentioned several stages of the process of electronization, the process itself is one of the toughest tasks faced by any local government. These stages are part of information security in the mean of conducting the local administrative policy.

In order to determine the status of e-services in the local administration, a representative survey has been conducted. The survey was conducted in early 2012 in 55 municipalities in the Republic of Bulgaria (NSI) [6] using a questionnaire consisting of six main sections, namely (Fig. 3):

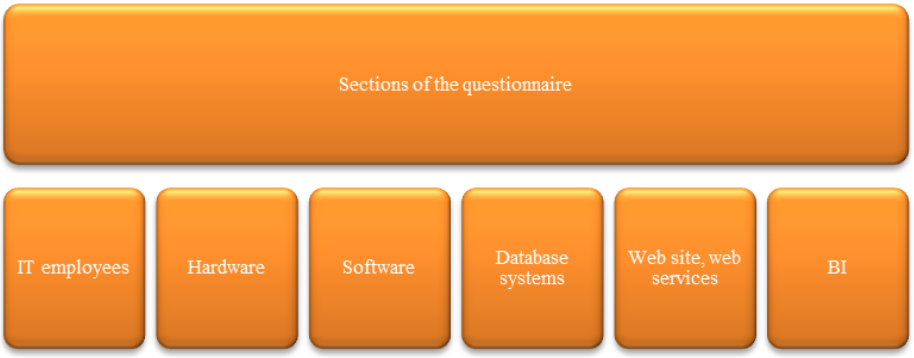


Fig. 3. Structure of the questionnaire

The territorial distribution of the municipalities in the sample is shown in Fig. 4.

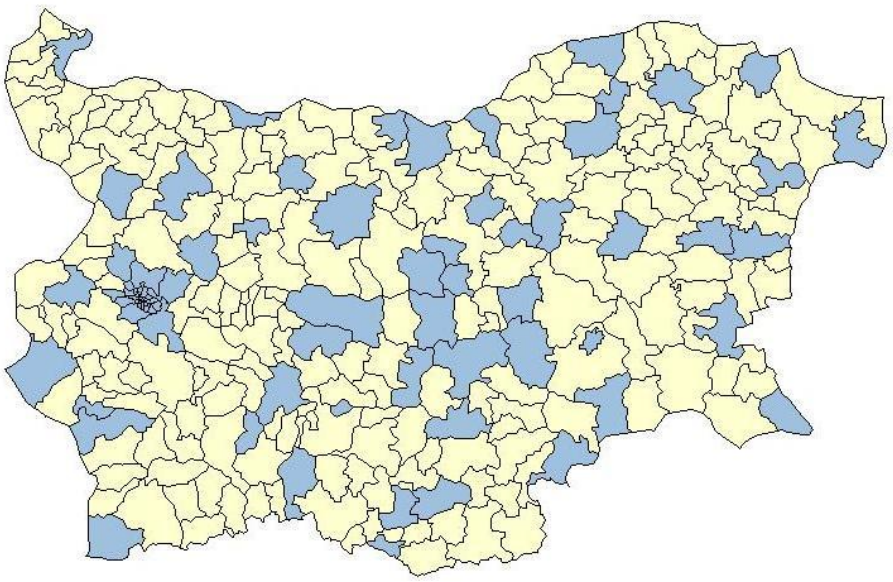


Fig. 4. Territorial distribution of the municipalities in the sample

The results of the responses to the question "Is it possible through the website of your municipality any online services to be performed?" are presented in Table 1.

Table 5. Online services

Answer	Municipalities
Yes	**34**
No	**21**

The results of the responses to this question show that more than half of municipalities provide online services - 34 municipalities. Unfortunately, the larger is the percentage of municipalities that do not provide such services - 40%. In order the electronic local government to operate, the web sites of the municipalities have to provide various electronic services.

The results of the responses to the question "Does your municipality provide some web services for connection with other systems?" are presented in Table 2.

Table 2. Web services for connection with other systems

Answer	Municipalities
Yes	**17**
No	**38**

The development of e-local government requires a connection to each municipal computer system with external systems and databases. Practice shows that this principle of data sharing is implemented through web services. Data from surveyed municipalities shows that such data transfer is possible only in 17 of all 55 municipalities. In the remaining 38 municipalities there is no existing software connection to external systems.

Summary

Empirical research and analysis made give us a reason make a formulation of a number of conclusions:

- Relatively rarely a single municipality hires new IT specialists;
- Municipalities have good hardware resources;
- A small number of local administrations have the opportunities for online services;
- More of the municipalities do not have systems for early warning systems;

These findings give us grounds for taking appropriate steps towards creating conditions for the expansion of electronic services at the local level.

References

1. Updated decentralization strategy 2006 – 2015
2. European Charter of Local Self-Government
3. Law on Local Self-Government and Local Administration
4. Municipal Budget Act
5. http://www.lex.bg
6. http://www.nsi.bg

Methodological issues of dynamising database structure in municipalities

Plamen Milev

University of National and World Economy – Sofia

Abstract. This paper focuses on the structure of municipal documents. These documents are official applications, which provide any specific facts and information in a particular type of law. In this sense, each local administration has its own document structure, described in the existing municipal regulations. Under the structure of municipal documents we understand the list of attributes of the document and their meanings in relation to this document. Typically, this structure does not change during the calendar year, but in many cases the structure is different for any of the following or even for each subsequent calendar year. Therefore, the paper examines some methodological issues of dynamising database structure in municipalities. For this purpose, five software methods are discussed – software method for defining documents, software method for presenting documents as metadata, software method for saving documents presented as metadata, software method for storing documents as metadata, software method for retrieving documents saved as metadata. The paper ends with conclusions about the advantages of using a software methodology for dynamising database structure in municipalities.

Keywords: Methodology, database, software method.

1 Software methodology for dynamising database structure in municipalities

For the purpose of establishing a software methodology for dynamising database structure in municipalities we choose a system of software methods that are subordinate to each other i.e. system of consistently associated methods. In this sense, by "software methodology" will be understood set of software methods and rules of interaction between them (fig. 1):

- Set of software methods. In terms of achieving the dynamic structure of municipal documents we offer software methodology consisting of five software methods, namely: software method for defining documents, software method for presenting documents as metadata, software method for saving documents presented as metadata, software method for storing documents as metadata, software method for retrieving documents saved as metadata;
- Rules of interaction. The basic rule in the present study is the relationship between software methods that are part of the proposed methodology for dynamising database structure in municipalities – the result of the execution of any method within the methodology is a gateway to the next method in the methodology (the result of the implementation of the last method within the methodology is relevant

in relation to the functionality of the first method) and the existence of any method makes sense only in the formation of the completion of the software methodology.

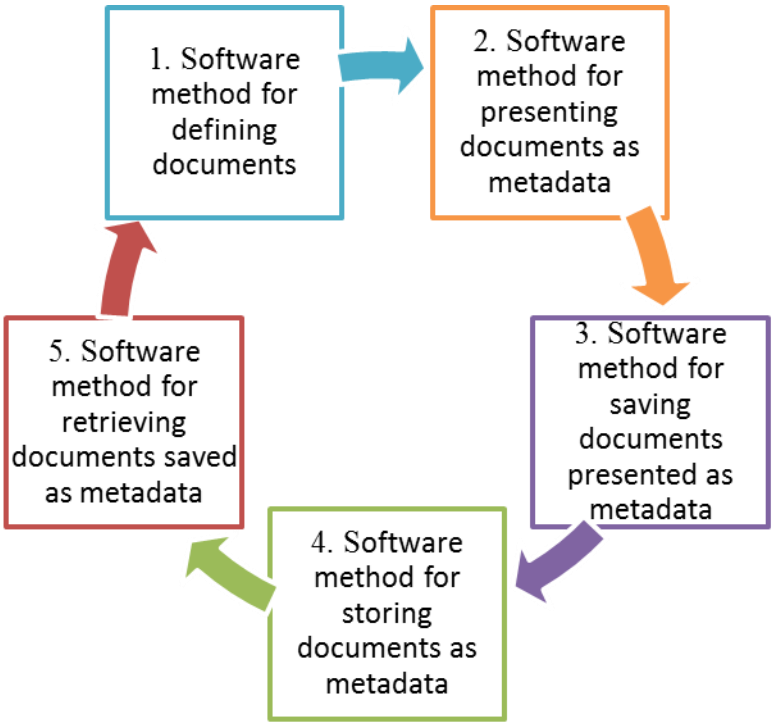


Fig. 1 Relation of software methods within the software methodology

In terms of research and in terms of the proposed methodology for dynamising database structure in municipalities, for us documents are forms that provide some concrete facts and evidence in a particular order. In this sense, each document has its own structure. Under the structure of the documents we mean the list of attributes for each document and their meaning in relation to this document.

From the point of view as a software solution, this set of five software methods, that form the software methodology for dynamising database structure in municipalities, represents a software environment that provides opportunities for initial definition of documents and subsequent use of the instances of these documents. This environment includes means of saving, storing and retrieving documents. Under "software environment" in the sense of this formulation, we understand software framework, which will make it possible for the set of the software methods to achieve their functional targets.

2 Software methods

The first method within the set of software methods, that build the software methodology for dynamising database structure in municipalities, is a **software method for defining**

documents. This method provides an opportunity for defining municipal documents, namely:

- Defining basic characteristics of the document, which includes the name, purpose, validity and other basic features;
- Defines the structure of a document that includes a list of attributes of the document and their meaning in this document.

Software method for defining document aims to create a universal algorithm for defining documents with different structures in which similar documents or different versions of a document can be created using inheritance based on already defined document.

Graphic description of the method for defining documents is shown at fig. 2.

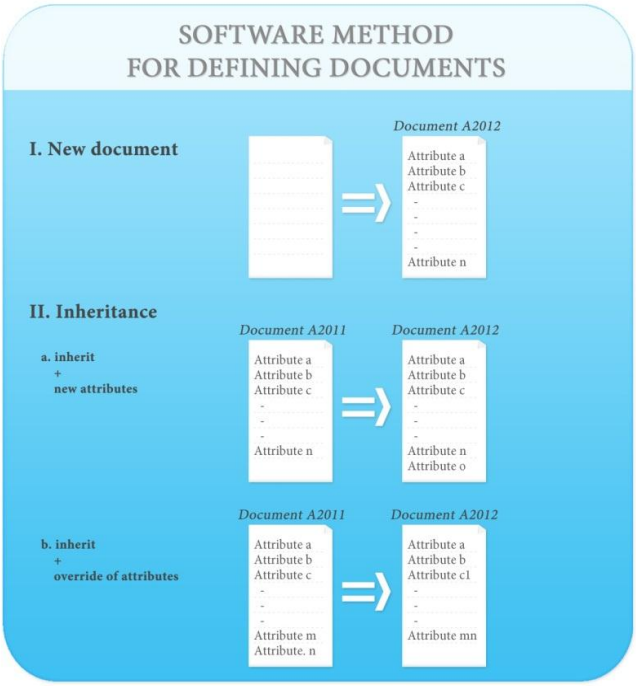


Fig. 2 Graphic description of software method for defining documents

The result of the implementation of the software method for defining documents represents a model of definition of document that includes basic features and a list of attributes (new and/or inherited).

The second method within the set of software methods, that build the software methodology for dynamising database structure in municipalities, is a **software method for presenting documents as metadata**. Under the term "metadata" within the formulation of

the method as a software method for presenting documents as metadata, we understand the description of the details of the documents in the form of data with a higher level of abstraction. By "higher level of abstraction" we describe the possibility of using metadata stored in different ways (database, index, file structures) through a unified approach for saving, storing and retrieving data. In this sense, the second method within the software methodology provides a unified scheme for presenting the attributes and their values of the defined documents as metadata with the necessary level of abstraction, namely:

- Defining basic characteristics of the document, which includes the name, purpose, validity and other basic features;
- Describing the copies of the documents in the form of metadata, this includes a unique identifier of the copy of the document and a list of values for the attributes of this document.

Software method for presenting documents as metadata aims to establish a universal scheme for describing documents with different structures where on the one hand we describe the definitions of documents and on the other - their copies of the principle attribute – value.

Graphic description of the method for presenting documents as metadata is shown at fig. 3.

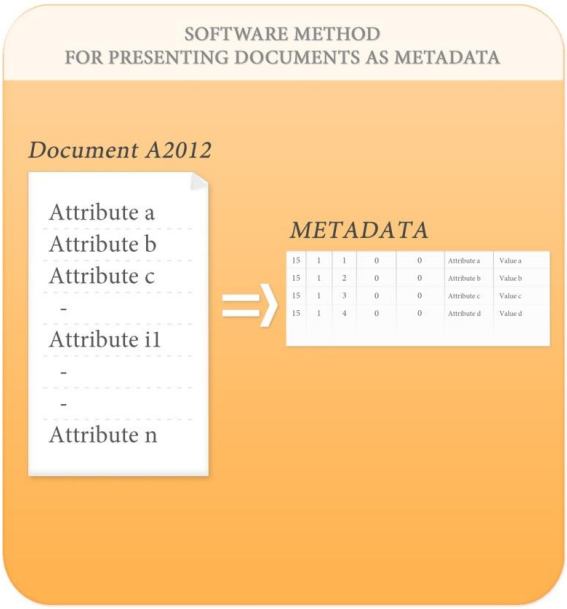


Fig. 3 Graphic description of software method for presenting documents as metadata

The result of the implementation of the software method for defining documents represents a model of presentation of documents in the form of metadata, which includes the definitions of documents and their copies.

The third method within the set of software methods, that build the software methodology for dynamising database structure in municipalities, is a **software method for saving documents presented as metadata**. This method provides an opportunity for saving documents as metadata, namely:

- Metadata, which represent definitions of documents, this includes the ability for saving in the form of SQL scripts, index data and data array;
- Metadata, which represent copies of the documents, this includes the ability for saving in the form of SQL scripts, index data and data array.

Software method for saving documents presented as metadata aims to convert the metadata with the described definitions of documents and their copies into an appropriate form for saving the metadata in some way.

Graphic description of the software method for saving documents presented as metadata is shown at fig. 4.

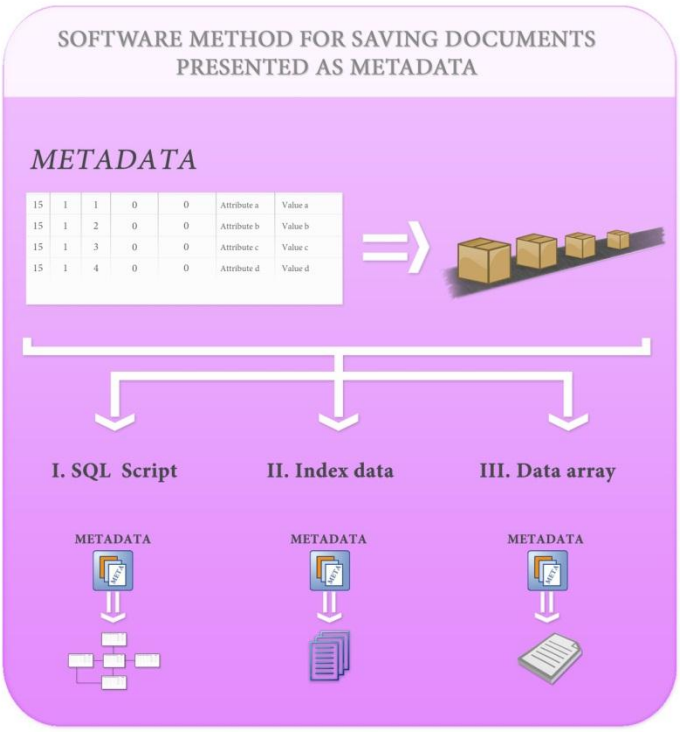


Fig. 4 Graphic description of software method for saving documents presented as metadata

The result of the implementation of the software method for saving documents presented as metadata represents SQL script, index data or data array, which include the content of the metadata.

The fourth method within the set of software methods, that build the software methodology for dynamising database structure in municipalities, is a **software method for storing documents as metadata.** This method provides the ability for storing metadata, namely:

- Metadata, which represent definitions of documents, which includes the ability for storing in a relational database, index and file structure;
- Metadata, which represent copies of documents, including the ability for storing in a relational database, index and file structure.

Software method for storing documents as metadata aims to maintain the stored definitions of documents and their copies in the appropriate manner for subsequent retrieval.

Graphic description of the software method for storing documents as metadata is shown at fig. 5.

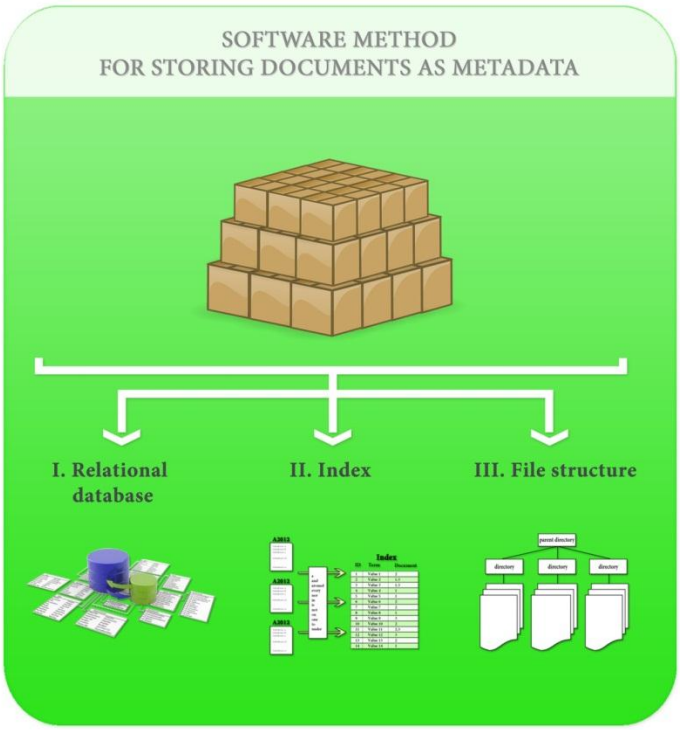


Fig. 5 Graphic description of software method for storing documents as metadata

The result of the implementation of the software method for storing documents as metadata represents stored metadata in the form of tables in a relational database, indexed data in an index or data array in a file structure.

The last method within the set of software methods, that build the software methodology for dynamising database structure in municipalities, is a **software method for retrieving**

documents stored as metadata. This method provides an opportunity to extract metadata, namely:

- Metadata, which represent definitions of documents, this includes the ability for retrieving metadata from SQL script, index and data array;
- Metadata, which represent copies of the documents, this includes the ability for retrieving metadata from SQL script, index and data array.

Software method for retrieving documents stored as metadata aims to convert the responding SQL scripts, index data and data arrays, which represent definitions of documents and their copies, into the metadata model used by the second method within the software methodology.

Graphic description of the software method for retrieving documents stored as metadata is shown at fig. 6.

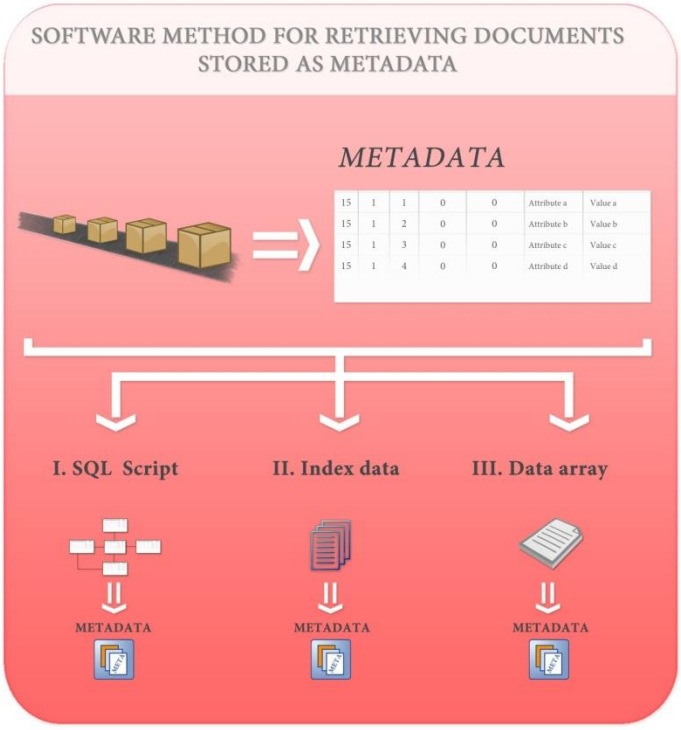


Fig. 6 Graphic description of software method for retrieving documents stored as metadata

The result of the implementation of the software method for retrieving documents stored as metadata represents a model of presentation of documents in the form of metadata, which includes description of definitions of documents and their copies, extracted from relevant metadata repository – a relational database, index or file structure.

Development of methodology for measuring the economic impact of ERP systems in small and medium enterprises

Natalia Futekova
University of National and World Economy

Abstract. Development and implementation of ERP systems is a difficult and responsible task. The success of future work with the system depends largely on the successful selection of modules for deployment and on the course of the process. Internationally, the theory and practice of these systems doesn't provide a uniform approach for determining the consistency of implementation of the various possible components of the system and even less a specific measure of the success of this process. For these reasons, the report examines options for developing a methodology for measuring the economic impact of ERP systems in small and medium enterprises.

Keywords: methodology, ERP, SME.

1 Introduction

Enterprise resource planning systems (ERP systems) belong to the class of business software systems that are characterized by a high degree of complexity, the design of which is based on best business practices. These systems are built on a modular basis with a common database and a large scale to cover the whole factory and perform supply management, production, sales, finance and accounting, etc.

Implementing an ERP system is a difficult and responsible task. The successful selection of modules for implementation and the conduct of the process depend largely on the success of future work with the system. Internationally, the theory and practice of these systems provides a unified approach to consistency in implementation of various possible modules of the system, much less a concrete measure of the success of this process. Moreover, over the world and in also in Bulgaria unsuccessful attempts to implement ERP systems are reported, followed by a serious loss of time, investment and customer interest. It is well known that much of the procedure of implementation is highly dependent on the specific business processes in enterprises. They in turn are determined by the economic environment of operation, the characteristics of the industry and the industry of the qualifications, skills and competencies of employees, and other macroeconomic conditions. To this variety of factor conditions should be added and also a strictly specific organization in the business of a company, which may be an additional factor complicating the implementation process.

Overall, there is considerable interest in the issues of identifying the factors contributing to the successful implementation of ERP systems. This is even more concerned about the problem of measuring and evaluating the effectiveness of the implementation process, and the efficiency gains resulting from the implementation.

2 Implementation of ERP systems

Implementation of ERP systems may be defined as a process within that definition of "process" in Bulgarian dictionary [1], namely as "a set of successive steps, a series of events, changing conditions or stages of something". In the scientific literature, process management is called "management process" and it defines the process as a strategic asset of any organization. On the other hand the management of the implementation process can be associated with certain of P. A. Smart order management process – "management approach of bringing all aspects and activities of an organization to the needs of service users"[3]. This effectively means that the implementation of an ERP system, allowing it to be controlled and directed in a certain direction (Fig. 1).

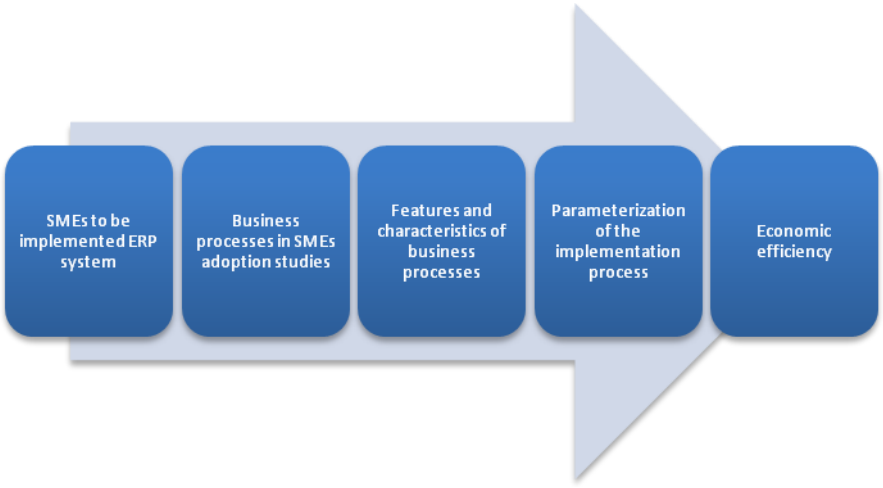


Fig. 1 Features of deployment

The determination of these characteristics of the analyzed business processes requires a thorough investigation:

- Information needs;
- Information flows;
- Applicable regulations;
- Documents used;
- Used software;
- Qualifications and competencies of the employees;
- Characteristics of the management company;
- Characteristics of planning and forecasting.

We believe that all of these activities allow the definition of indicators to measure the effectiveness of implementation of ERP system alleged or its individual modules. On the other hand, as seen above, implementation process is also complex combination of activities and more recently the sequence of activities. It is quite possible, according to the information needs of survey, part of this sequence to be changed or modified.

In this paper, we consider the possibility that the two parts of the research process are connected to each other and determine the usefulness and impact of deployment. This makes possible the following relation (Fig. 2)

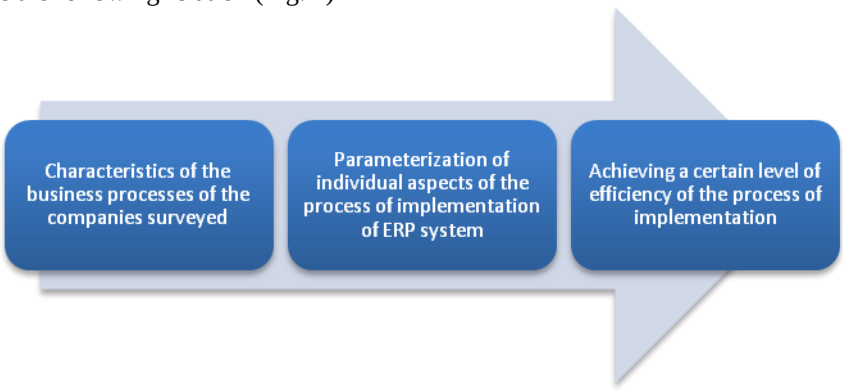


Fig. 2 Influence of parameterization on the implementation of the achieved level of performance

The results of the development of this problem can find specific practical applications, such as: the development and application of a particular software product (tool), which even at the stage of research on the characteristics of the processes studied to be a firm proposal of modification and parameterization of certain aspects of the innovatory process and anticipate achieving a level of performance. This will allow an "effective design" at the stage of investigation. At the same time, this topic is related to current practices in the implementation of ERP systems, such as those used by the world leading company Microsoft [2]. The main problem is finding oriented approach to develop a methodology to measure the effectiveness of the alleged deployment of the ERP system at the stage of investigation. Regarding the implementation of ERP systems, we can identify the following important points, namely:

- Performing a comprehensive empirical survey of SMEs implemented or in the process of implementation of ERP system will allow clarification of the individual characteristics of the analyzed business processes and allow the definition of indicators for different methods of methodology for measuring the effectiveness of implementation. But we need to recognize that there are a number of features for each individual company in which the synthesis process is smooth with these differences;
- The process of implementation of any software in itself is dependent on software implementations (not in any ERP system endless parameterization of this process can be made);
- Measuring the assumed level of efficiency of the implementation of an ERP system in a company will perform the exploratory stage. At the same time the implementation is over a certain period during which the analyzed company continues to evolve and change, which can alter perceptions of effectiveness. It is therefore necessary to design a specific component (method) of the assessment methodology that reflects this process;

- The availability of qualitative and quantitative characteristics in different parts of the business processes in SMEs. Measurement of any effects on quality characteristics are particularly difficult and requires a scheme for converting qualitative into quantitative values (to be given meaning and effect to measure);
- Restrictive condition is the availability of budget for implementation. This raises the following important question: "What is the cost of parameterization of the elements of the implementation in order to achieve higher efficiency of software?" Our view is that there is a correlation between the effectiveness of implementation and cost-effectiveness.

2 Development of methodology for measuring the economic impact of implementing ERP systems

Based on these and other aspects of the research problem, we propose a methodology for measuring the economic impact of implementing ERP systems. Fig. 3 presents a general scheme for the application of the methodology for measuring effectiveness.

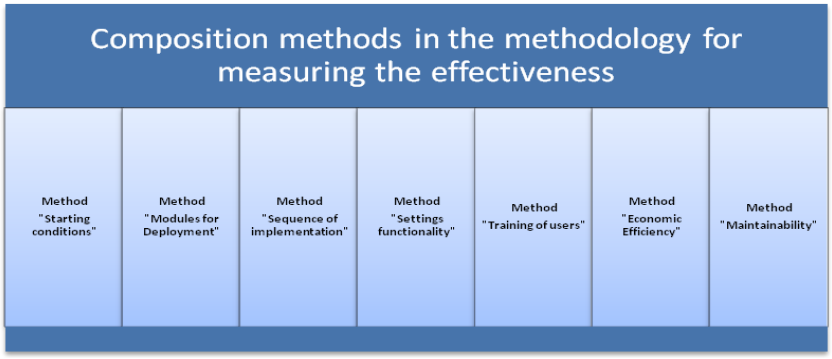


Fig. 3 General concept of methodological decision

General description of the methodological solution is:

1. The exploration phase in the implementation of an ERP system defines the first idea for a series of stages in the implementation and characteristics of different stages.
2. Indicators in the methods of the proposed methodology are an analysis of the activities of study in the small and medium enterprises.
3. Defining the initial value of the effectiveness of implementation based on the research done and the original idea of the process of implementation.
4. Defining requirements to identify opportunities for increased efficiency.
5. Application of a software prototype to identify opportunities, means and ways to improve the effectiveness of implementation through adaptation and parameterization of individual elements of the deployment.
6. Final identification of the opportunities for parameterization of the elements of the implementation process.

7. Parameterization the process of implementation and determining the ultimate cost effectiveness as a result of adaptation.

Each of the methods for evaluating the effectiveness of the composition provides innovative technology development:

- General description of the method including the type and presentation of the overall performance of the method and its place in the proposed methodology;
- Define a list of indicators in the composition of each method;
- Development of an algorithm to determine the ultimate cost effectiveness as measured by that method. Prerequisite and part of the algorithm is presented by mathematical means;
- General graphics scheme description of each method in the composition of innovative technology.

For each of the proposed methods for evaluating the effectiveness we propose the following scheme (Fig. 4).

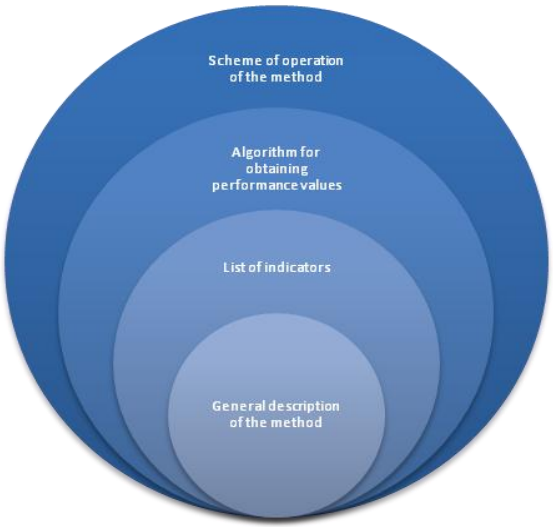


Fig. 4 Scheme of operation of each method in the composition of the methodology

We have described the general concept of the nature and composition of the methodology for measuring the effectiveness. The description made gives us grounds to believe that it can be successfully implemented in the form of software.

3 Conclusion

In conclusion, it should be noted that at present, there are no known analogues, adapted to the Bulgarian market conditions in the area of such technology software solutions. It can be considered that the development resulting from the application of the methodology of the final product will actually lead to improves in the competitiveness of many companies from different economic sectors. This will be possible by allowing the introduction of advanced software solutions.

References

1.Andreychin, L., Et al. Bulgarian dictionary, "Science and Art", Sofia, 2008
2.Microsoft Official Training Materials for Microsoft Dynamics ® Project Managing Microsoft Dynamics ® AX, Microsoft Dynamics GP, and Microsoft Dynamics NAV Implementations with Microsoft Dynamics SureStep 2010
3.Smart, P.A, Maddern, H. & Maull, R. S. (2008) Understanding Business Process Management: implications for theory and practice, British Journal of Management, 2008

Model of intelligent system for user activity monitoring with dynamic behavior profiles

Anton Palazov

Abstract: The possibilities for the application of the methods for data mining in the creation of the profiles of users of information systems to monitor their activiteis in order to protect the data from the internal threats and dynamic processing of their authorization in respect of a variety of information resources and services are analyzed. A limited number of typical elements of user activity within the framework of the information system of the organization are defined, which can be measured using data from log-repositories. Formed values are processed in static by algorithms for Cluster analysis and groups of users according to their typical activity in the middle of the information system of the organization are defined.

Introduction

The purpose of this publication is to present current results in the study of possibilities for the application of the methods for Data Mining in the field of information security, and in particular to form profiles of users of information systems, with which to monitor their activity to counter internal threats and dynamically improving their authorization in respect of a variety of resources and services.

Planned in the framework of the research experiments are carried out with Rapid Miner – one of quality and freely available software products for implementing methods of Data Mining. The overall structure of the basic process of dynamic build profiles of behavior for monitoring user activity in information systems is shown in Fig. 1.

1. Input Data for testing the Model

Experimental model for monitoring user activity and dynamically build profiles of behavior is based on the information on the following typical user actions and their duration:

- typically executed programs;
- typically accessed folders which are organised for use by multiple users within the organization;
- commonly used Internet resources.

The input data that are used in the creation and testing of the model to build profiles of user activity, consist of 200 and activities are organized in the file LogIN, which is as follows:

- user name that is registered and recognized by the system;
- date of occurrence of the analysed event;
- time of occurrence of the analysed event;
- type of audited object – application, data folder, Internet resource;

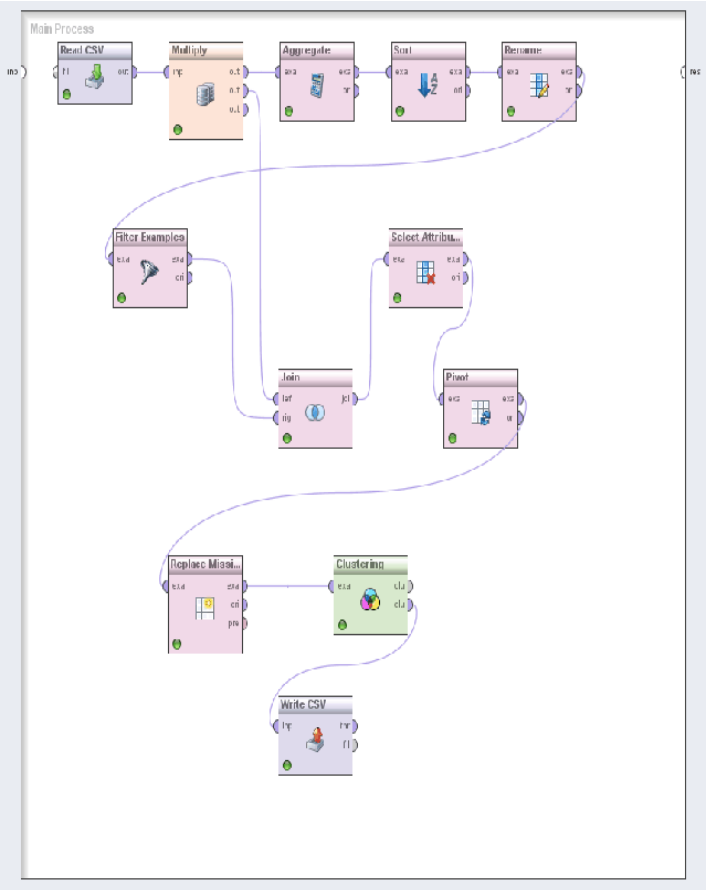


Fig.1 Model of basic process for behavior profiles creation

- activated object – name of executed application, path to an opened folder with data for shared use, URL address of Internet resource;

- duration of the action in seconds.

```
apal,20120729,23:44,app,MSOffice,361
usr1,20120729,23:54,app,MSNav,248
usr2,20120729,23:55,dat,\\srv1\shFolder\,104
usr3,20120729,12:15,url,www.url1.com,960
usr3,20120730,23:44,app,MSNav,36
usr1,20120731,23:54,url,www.dir.bg,24
usr2,20120801,13:45,dat,\\srv1\shFolder\,104
usr3,20120801,12:48,url,www.url2.com,240
apal,20120803,06:48,app,MSOffice,1012
usr2,20120804,23:55,dat,\\srv1\CommonFolder\,216
usr3,20120805,12:15,dat,\\srv2\EmailRepository,240
apal,20120805,13:44,app,MSNav,1612
usr7,20120806,11:51,app,Outlook,124
usr7,20120729,23:55,dat,\\srv1\shFolder\,104
```

Fig.2 Input Data for Model Testing

Part of the contents of the input data used in testing of the model is shown in Fig. 2.

2. Extracting profile elements

The model provides dynamic modification of the elements which make up the profiles of user activity. In the set of such elements are included applications with the most prolonged usage within a specific organization (determined by the accumulation of the duration of the work with them from registered users), and shared folders and Web resources that are most frequently accessed (defined as the number of the resource accessing according to the data from the log file).

Obtaining these results is achieved by the following procedure:

- input data is aggregated by type of object (application, folder, or website), such as duration of use of the programmes is summed, and the openings of folders and pages are counted;
- the results obtained are arranged in descending order of the aggregate score for the frequency of use of the resources;
- this part of them, which has decisive significance in shaping the profile of user activity is filtered (extracted). In realization of this model as a profile forming are chosen these activities, whose aggregate (total) score exceeds a specified threshold of 1,500 units. In accordance with this threshold profiles of behaviour will be formed by their meanings on 7 indicators, of which 2 applications, 3 folders with data for common use and 2 Internet resources.

The results obtained by processing the test data for the model are shown in Fig. 3.

Row No	ObjectName	TotalDuration	TotalCount
1	MSNav	19151.0	38.0
2	MSOffice	12104.0	29.0
3	\\srv1\shFolder\	3378.0	40.0
4	Outlook	3042.0	15.0
5	\\srv1\CommonFolder\	2752.0	12.0
6	www.url1.com	2648.0	26.0
7	\\srv2\EmailRepository	2236.0	16.0

Fig. 3 Elements forming the profile of behavior, and their estimations

3. Forming profiles of user activity

To build profiles of user activity within the study was chosen one of the methods of Cluster Analysis – k-Means. The implementation of this method for Data Mining requires input data can be prepared in a certain way. To meet these requirements the following actions are performed:

- join the get input from log-file with the set of retrieved elements to form profiles of behaviour according to the frequency of their use, allowing each user to receive an individual assessment on these indicators;

Row No	User Name	Total Duration MSNav	Total Duration MSOffice	Total Duration Outlook	Total Duration www.url1.com	Total Duration \\srv1\Common Folder	Total Duration \\srv1\ shFolder	Total Duration \\srv2\Email Repository
1	apal	2658.0	6912.0	124.0	96.0	1.0	1.0	1.0
2	usr1	7910.0	216.0	144.0	1.0	1.0	1.0	1.0
3	usr2	36.0	216.0	124.0	1.0	864.0	2480.0	1.0
4	usr3	345.0	447.0	460.0	2304.0	1.0	1.0	1804.0
5	usr4	1.0	4313.0	1.0	1.0	1.0	378.0	378.0
6	usr5	3532.0	1.0	50.0	222.0	1.0	1.0	1.0
7	usr6	4670.0	1.0	1.0	26.0	1.0	1.0	1.0
8	usr7	1.0	1.0	2140.0	1.0	1888.0	520.0	54.0

Fig. 4 Distribution of profiling items by users

- a Pivot Example Set, which summarized the personal activities of each of the registered users of the information system for each of the identified elements of their profiles of behaviour – duration of use, frequency of access to shared folders for information resources, the number of views of Internet resources. The contents of this relation for the model test data is shown in Fig. 4. Before creating the distribution, missing values are replaced with 1, which is a requirement to use the chosen method of Cluster analysis.;
- so formed data are processed by the k-Means method for Cluster Analysis which groups users according to their activity in terms of the extracted behavior profiles forming elements. The method is used in the suggestion for the creation of 3 groups and performance of 100 iterations for optimal grouping.

ExampleSet (8 examples, 2 special attributes, 8 regular attributes) View Filter (

Row No.	id	cluster	UserName	TotalDurati...	TotalDurati...	TotalDurati...	TotalDurati...	TotalDurati...	TotalDurati...	TotalDurati...
1	1	cluster_0	apal	463	6912	124	96	1	1	1
2	2	cluster_1	usr1	7910	216	144	1	1	1	1
3	3	cluster_1	usr2	36	216	124	1	864	2480	1
4	4	cluster_0	usr3	345	447	460	2304	1	1	1804
5	5	cluster_1	usr4	1	4313	1	1	1	378	378
6	6	cluster_2	usr5	3202	1	50	222	1	1	1
7	7	cluster_2	usr6	4670	1	1	26	1	1	1
8	8	cluster_2	usr7	1	1	2140	1	1888	520	54

Fig.5 Cluster Analysis results

The results for formed groups, breakdown of users between them and the specific characteristics of each group are shown in Fig. 5. It shows that users are pooled into three groups containing respectively 2, 3 and 3 members with typical activities, such as using Office applications, working with ERP application of the organization and implementation of communication activities.

4. Conclusion

Dynamically formed by means of Cluster analysis in Rapid Miner profiles of behavior can be used by system administrators to modify the permissions granted to the users of the information system in terms of available information resources. This will allow, at this stage, the functions of monitoring of user activity and detecting anormalies (exceptions) from ordinary activities to be delegated to the operating system or its appendices in larger organizations, such as the security information and event management (SIEM).

References

1. Rapid Miner, User manual, Rapid-I, 2010.
2. P. Tryfos, Cluster Analysis, 1997.
3. A. Lane, Fundamentals of User Activity Monitoring, InformationWeek Reports, 2012.
4. www.dataminingtools.net/rapidminer

Cost - Benefit analysis as a managerial aspect of Business Intelligence implementation

Ivan Andonov

Abstract. This research is to point out the issues faced in the Cost – Benefit analysis preparation and to offer possible solutions of the mentioned issues. This article describes the most important stages of BI implementation process and what their contribution is in making of Cost- Benefit analysis. The key element here is how to predict something that up to now is unknown i.e. how to measure and calculate intangible information and how to prove in the Cost – Benefit analysis that is a beneficial decision to be implemented Business Intelligence platform when there are so many unknown variables.

Keywords: Cost – Benefit analysis, Business Intelligence implementation process, Business Intelligence's issues.

This research is based on information provided by one of the leading companies in the energy sector of Bulgaria about this what the essential of Cost – Benefit analysis is in order to be taken in mind as a managerial aspect when we are talking about Business Intelligence implementation. The research aims to point out what the issues are when the Cost – Benefit analysis is prepared.

Here in this middle research, under Business Intelligence and Cost- Benefit analysis, the author means the following definitions:

Business Intelligence [1]:

This definition is given by the expert of Gartner Group and it is:" an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance."

Cost – Benefit analysis [2]:

A process by which you weigh expected costs against expected benefits to determine the best (or most profitable) course of action.

The preparation of this Cost – Benefit analysis (CBA) is from a huge importance. The reason for this is by preparing it the Board of Directors has a document proving how much this project of Business Intelligence platform implementation will cost. BI implementation should be described phase by phase and Dollar per Dollar. BI implementation is and investment and the most important for any investor is the Return on investment (ROI). CBA should involve all costs and revenues. Of course, it should be taken in mind in the beginning of the BI implementation some costs and benefits are unknown.

When we are going to implement the BI tool, we should know the steps or phases that we are going through. Some of the popular methodologies of leading companies are: figure 1 IBM BI methodology and figure 2:

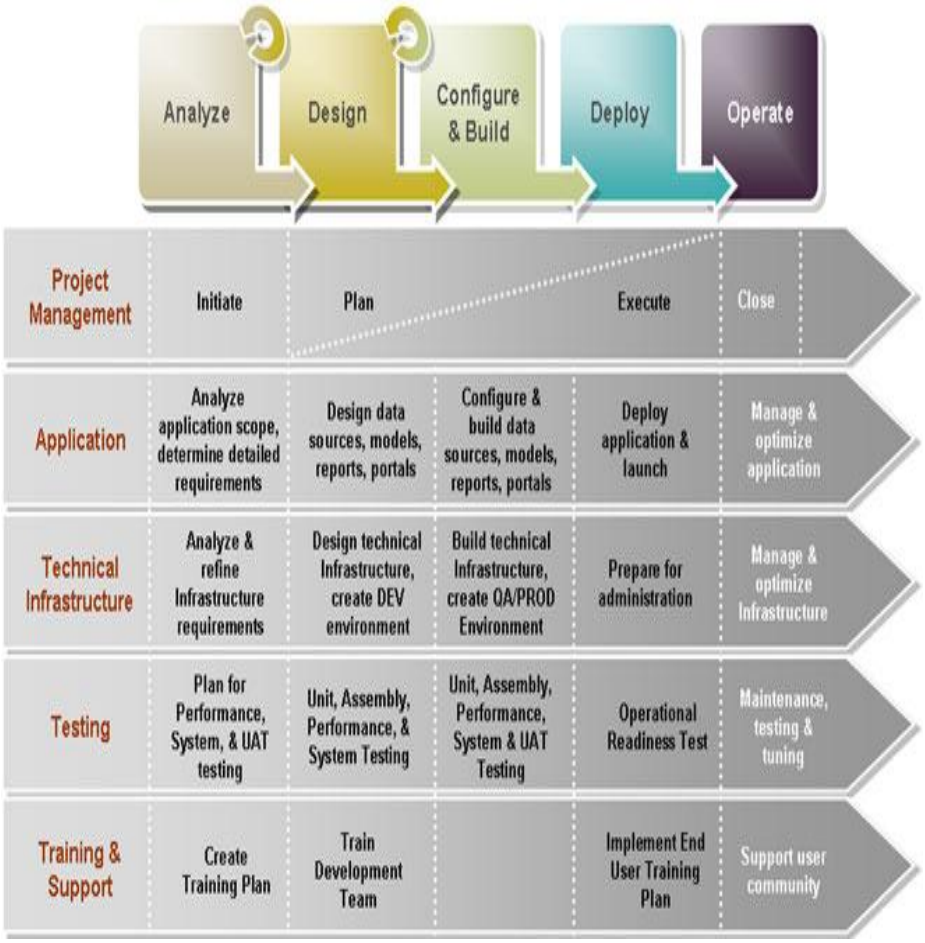


Fig. 1. IBM BI methodology [3]

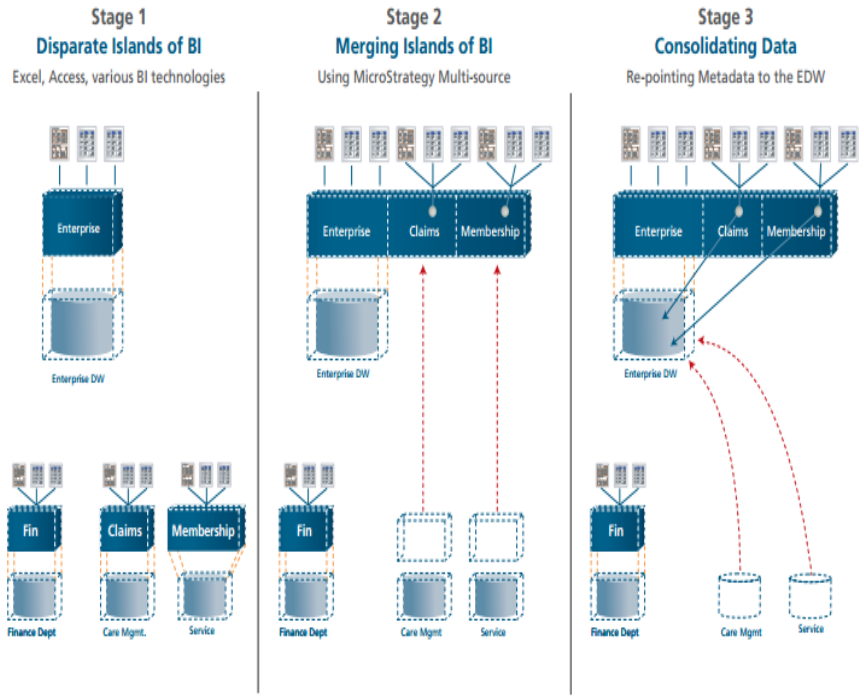


Fig. 2. MicroStrategy methodology for HP project [4]

The methodologies are key elements of BI implementation because they show step by step everything. So, it means the followed methodology shows and what costs will be and what benefits will appear and how to be calculated.

The main parts of every methodology are:

- BI project planning activities;
- Set goals;
- Cost – Benefit analysis of the BI implementation;
- To be defined Benefits of BI vs. changes in business processes and to be calculated;
- Platform and vendor selection;
- How data should be prepared
- Employees' training

In this context, it is clear that CBA should be well prepared and everything to be taken in mind. Every step is related to be spent money and this should be in CBA. The famous scientist and researcher Thomas Groh [5] invented the following table to present how costs of BI implementation should be measured and calculated.

Table 1 Determining External and Internal Labor Costs

	Internal			External		
Required Roles (initial)	Hrs.	Cost Rate	Total Cost	Hrs.	Cost Rate	Total Cost
Project Manager		$ –	$ –		$ –	$ –
Business Analyst(s)		$ –	$ –		$ –	$ –
Technical Lead/Solution Architect		$ –	$ –		$ –	$ –
Data Modeler(s)		$ –	$ –		$ –	$ –
ETL/Database Specialist(s)		$ –	$ –		$ –	$ –
Front-End Tool Specialist(s)		$ –	$ –		$ –	$ –
Stakeholders/Bus. Subj.-Matter Expert(s)		$ –	$ –		$ –	$ –
IT - Network/Administrator(s)		$ –	$ –		$ –	$ –
IT - System/DB Administrator(s)		$ –	$ –		$ –	$ –
Subtotal Labor Costs			$ –			$ –
Total						$ –

In the preparation of Cost – Benefit Analysis, it is not the issue how to be calculated the costs; the problem is how to be measured and calculated the benefits. The benefits of BI are the data of the enterprise that will be analyzed by using this BI tool and in this data is value, the real benefit for the company, but this could be known when the BI implementation process is complete. The real challenge is to be measured these benefits in the beginning of the BI process. The most difficult question to be answered is: How to measure the qualitative benefits? It involves:

- Improved information dissemination;
- Improved information access;
- Improved data quality and feedback from operational systems;
- Improved collaboration between business units.

Why it is a challenge to prepare a CBA that includes all effects of BI implementation? The answer is:

- Calculating the value of the BI solution is difficult due to the fact that it equals to calculating of the value of the potential information which is unknown;
- Intangible benefits;
- Real value of the information and its amount will be determined only after the investment is made.

Because of these issues it is recommended to be done the following:

- BI implementation project to be divided into smaller cascading stages where every next stage depended on the successful implementation and valuable payback of the previous stage;
- Sensitivity analysis;
- Development of new indicators how to measure this improved information.

When the whole process of BI implementation is described in details then it is going to be easy to be calculated the following key investment indicators:

- Return on investment (ROI);
- Internal rate of return (IRR);
- Net Present Value (NPV);

In this research in the area of BI implementation, the solving of the mentioned issue about Cost – Benefit Analysis is not a battle that could be won easily. The truth is that it should be compared CBA of this company to CBA's in other companies in the same area of business and to be used the things that they have learned as a result of BI implementation how CBA should be done in an adequate way.

Future research

The purpose of this research is when it is completed:

- To be invented indicators how this qualitative information to be measured;
- And to be done a comparison between other companies that implemented BI tool in the same area of business and to be pointed out the similarities and differences, the advantages of BI implementation in one of the companies and disadvantages of the other.

Experience is a key element of successful implementation of BI and an adequate prepared Cost – Benefit analysis.

5 Conclusion

Business intelligence is a high – tech tool that beneficiates a lot the company that uses it, but the issues that should be taken in mind when we are talking about it are:

- how to predict the benefits of BI implementation and to point them out in the Cost – Benefit analysis;
- how to choose an adequate methodology that will guarantee us a successful Business Intelligence platform implementation- is it better to develop our own methodology or to choose a methodology of a leading company in this area as IBM for example;

Cost – Benefit analysis is a key element in the beginning stage of BI implementation i.e. when it is taken the decision to be or not to be implemented and developed this tool in the

enterprise. It is crucial to be known all the costs as: purchase of software and hardware, maintenance, labor costs and other costs. Because even is all the benefits are not known, the board of Directors will know the costs and how to proceed in order to get benefits of this BI tool.

CBA is an important document that helps a lot in decision making process to be implemented the BI tool. That is why it should be prepared by experts and to include all possible data about the project.

References

1.http://www.gartner.com/it-glossary/business-intelligence-bi/
2.http://www.entrepreneur.com/encyclopedia/term/82116.html
3.http://www-01.ibm.com/software/analytics/cognos/services/imp-method.html
4.http://www.microstrategy.com/download/files/solutions/ByIndustry/MicroStrategyandHPHealthcare SolutionsBrochure.pdf
5.http://www.information-management.com/issues/20040101/7921-1.

FORMULATION OF CRITERIA FOR COMPARISON AND ANALYSIS OF ANALYTICAL APPLICATIONS WHICH WORKS IN REAL TIME

Mihail Konchev,

UNWE, Dept. „Information technologies and communications"

Abstract. Modern companies collect and store vast amounts of data, which in recent years increased progressively. A successful approach to making effective management decisions is the ability to analyze business information in real time. Management needs accurate information about the company at a specific time, processes that occur in it, the trends that lead to improvement or deterioration of its market positioning. An information system must in real time to monitor and record various events and processes to provide systematic information collected to evaluate and analyze and prepare reports and forecasts for development. Means for solving these tasks are analytical applications running in real time. Through these managers have the opportunity in real time to visualize and work with analytical data and draw trends in their company. In the literature on various aspects and perspectives are considered benchmarks and analysis applications for multidimensional analysis.

The paper aims to summarize the main aspects in the analysis of On-Line Analytical Processing (OLAP) applications and display the main features of the analytical applications running in real time. Based on deduced characteristics are defined criteria for comparison and analysis. In the end of the paper are determined weightings for each criterion.

Keywords: OLAP, real time, criteria for comparison and analysis, weightings

1. INTRODUCTION

The development of information and communication technologies is a prerequisite for fusion of information technology and decision making, which is based on the availability of reliable and timely analytical information. Such information could be obtained from the analytical applications running in real time. According to a International Data Corporation (IDC) "information in the world doubles every two years. In 2011 the world will create 1,8 zettabytes data. By 2020, it will generate 50 times more data and 75 times to increase the number of "data warehouses", while IT personnel necessary to manage this information will increase by less than 1.5 times' [1].

Following the laws of market, dynamically changing business environment in the era of information technology, as external factors for companies and the trend of exponentially increasing volumes of business data, play a role at two levels: users of BI applications and companies offering business intelligent applications.

The impact of these two factors on users of BI applications is confirmed by a study conducted by Philip Russom - General Manager of Research at TDWI. He said - increasing volumes of information had DW companies to integrate planning and integration of new technologies and solutions for processing and analysis of data. As part of market mechanism - companies offering BIS also appear in a highly competitive environment. Confirmation of this is the matrix of Gartner intelligent business platforms, data for 2012 [2]. In a highly competitive environment, the choice of BI platform does not change the main reasons

justifying the decision to integrate the BI system, but create a scientific problem with the lack of generally reasonable criteria for evaluating individual applications for real-time analysis – RTOLAP.

2. MODELS FOR EVALUATION OF ONLINE ANALYTIC APPLICATIONS

In 1993 E.F. Codd published an article entitled, "Providing OLAP users - analysts as a" term "information technology." In this article, E.F 12 Codd formulated the basic principles of OLAP. During 1995 and adding six more. Codd divided into four groups, these rules, calling them "signs" (Table.1) [3].

SIGNS	FUNCTIONS
Key Features	• Multidimensional views of data • Intuitive handling data. • Accessibility: OLAP as a mediator. • Package for interpretive data mining. • Model for OLAP analysis. • "Architecture" client-server. • Transparency. • Versatility to consumers.
Special Features	• Processing of non-normalized data • Excluding missing values • Treatment of missing data • Save the results of OLAP. Keeping results separate from the original data
Specifics of reporting	• Flexible reporting • Standard Performance Reports • Automatic adjustment of physical layer
Management of the dimensions	• Versatility of dimensions • Unlimited dimensions and aggregation levels • Unlimited interactions between dimensions

Table.1: Codd's principles of OLAP

A well-designed data warehouse provides users with fast, easy to use and effective analytical capabilities for processing large datasets. The popular quick test for analysis of shared multidimensional information (FASMI) used to achieve this.

As proposed by Nigel Pendse and Richard Creeth FASMI valuation model, OLAP applications must be able to meet the analytical requirements set by the user, and be easy to use. Analytical applications for multidimensional data analysis should include relevant security requirements and to provide a multidimensional view of the data, including full support for multiple hierarchies. In addition, all data received from the data OLAP applications to be accessed by different applications [5].

1. **Fast response** - criteria on the performance of OLAP application is able to return a response to consumer demand within five seconds, in very rare cases, the response

time reaches 20 seconds. For simple queries necessary response time should not exceed one second.

2. **Analytical** - means that the system can cope with any business logic and statistical analysis that is relevant to the user, and is presented in a convenient format for it and view the resulting analytical data.
3. **Shared** - means that the system meets all requirements for security and confidentiality of data (perhaps to a level cell database) and, if necessary, introducing restrictive policy update certain level.
4. **Multidimensionality of the data** - this is a basic requirement for OLAP applications. If you need to select a single word for the definition of OLAP, this is - multidimensionality of data. The system must provide a multidimensional view of data, including full support for multiple hierarchies, as this is certainly the most logical method of analysis of business organizations.
5. **Information** - this is a collection of all necessary data and extracted information that is relevant to the analysis.

3. PROCESSES GOVERNING THE OPERATION OF RTOLAP APPLICATIONS

In terms of process, ensuring work Real-Time OLAP (RTOLAP) applications and ability to quickly retrieve, organize, summarize and present multidimensional data at any time when a change occurs in the main transactional data source can be summarized these processes:

- Notification of change in the relational database;
- Processing of metadata;
- Develop a plan for handling RTOLAP cube;
- processing of RTOLAP cube;
- Data processing, including read and write data;
- Generating aggregates and indexes;
- Updating RTOLAP cube;
- Processing RTOLAP elements;
- Construction schedule RTOLAP multidimensional cube;
- Processing RTOLAP dimensions (Dimensions);
- Treatment groups RTOLAP measures (Group of Measures);
- Treatment groups RTOLAP dimensions (Group of Dimensions);
- Processing of hierarchical structure RTOLAP dimensions;
- Processing of multidimensional levels RTOLAP structure;
- Execution of requests RTOLAP cube;
- Entry;
- Confirmation of successful completion process;
- Batch processing of Transact-SQL reports;
- Perform a query.

4. CRITERIA FOR COMPARISON AND ANALYSIS OF RTOLAP APPLICATIONS

Based on analyzing the main features of online analytical applications running in real time and the above processes governing the operation of these applications, described below author assessment model RTOLAP applications, the following criteria belonging to a group of sign "real-time" evaluation of the applications:

- Optimal caching of requests to the database;
- Ability to pre-aggregation of data;
- Time synchronization between client and OLAP OLAP server;
- Processing time building and deploying RTOLAP cube;
- Processing of metadata;
- Processing RTOLAP cube;
- Data processing;
- Generating aggregates and indexes;
- Updating RTOLAP cube;
- Processing RTOLAP elements;
- Execution of requests RTOLAP cube;
- Execute the query to SQL Server.

Based on the specific characteristics displayed two of the most famous in theory and applied in practice models to evaluate applications RTOLAP E.F.Codd model and FASMI model. The following signs and associated groups these criteria to the evaluation of the applications:

- Functionality;
- Integration with existing applications;
- Productivity and performance;
- Security;
- Consumption of system resource.

From the point of view the similarities of the two models (EFCodd and FASMI model)assessment RTOLAP applications described below author's model for assessing applications, the following groups of signs and adjoining criteria for the evaluation of the applications:

- Create RTOLAP cube;
- Scalability.

Specific criteria for the author's assessment model RTOLAP applications are:

- Real time (described above);
- Price.

5. CONCLUSION

The companies which are planning to become analytical capacity in competitive advantage, should have an appropriate model for comparison and evaluation of analytical software to industry-leading companies

The choice of a software tool for analysis depends on the objectives, tasks, data volume and other factors that determine companies as critical to their business. The success of any

organization is finding the optimum proportions - performance and price, weighted according to the strategies, goals and resources of the company.

Under current management, the length of the business cycle is extremely condensed and thus the availability of fast, reasonable and efficient process of decision-making is an important prerequisite to improve the competitiveness of companies. Managers need the right information at the right time and the right place. In this regard, there is a critical factor in the business today is the need for special attention to the management of building initiatives business intelligence systems.. Therefore, more and more organizations expressed their full support for the implementation of intelligent business systems and develop strategies for their development.

The flexibility and performance of the processing and analysis of large volumes of data in real-time characteristics of many software solutions developed by key suppliers. Online analytical applications, real-time, are strategic potential that companies should evaluate as a means to achieve sustainable advantage.

REFERENCES

1. IDC, http://www.emc.com/leadership/programs/digital-universe.htm;
2. Gartner Research Report: 2012 Business Intelligence Platform Maqic Quadrant http://www.qlikview.com/us/explore/resources/analyst-reports/gartner-research-report-2012-business-intelligence-platform-magic-quadrant
3. E.F. Codd, S.B. Codd , C.T. Salley, “Providing OLAP users - analysts as a" term "information technology. ”,1993, http://www.minet.uni-jena.de/dbis/lehre/ss2005/sem_dwh/lit/Cod93.pdf
4. Nigel Pendse, Richard Creeth: The OLAP report. Business Intelligence, UK,1995, ISBN 1-898085-34-X, pages 9-10

Performance Bottlenecks of an Educational Social Network Built with Elgg

Venko Andonov[1],

[1] University of National and World Economy
venkoandonov@gmail.com

Abstract. A growing number of universities around the world are trying to increase the engagement in the educational process and to improve the professional skills of their students by using the latest trends in the online social networking in a learning context. There are numerous alternative solutions, including, but not limited to: building their own private educational social networks; using popular social networking sites such as Facebook and Twitter to create a familiar online environment for discussions and sharing information. Taking the first approach has not been extensively studied yet, because it usually requires modifying and extending an existing open source application to meet the specific requirements – like emphasizing the educational context and ensuring the authenticity of the participants. We have evaluated how our implementation of the popular open source application Elgg handles the growth in three dimensions – users, content and operation complexity. The results show that its data model might lead to significant performance issues.

Keywords: educational social networks, data models, performance

1 Introduction

An educational social network (ESN), sometimes also called Social Learning Network or Social Learning Environment, is a type of online social network which provides a set of tools and services to enrich the learning process for all of the participants in a given educational environment [1, 2, 3]. For example, in the higher education, such a system might be an engaging way to promote some desired activities among the students, including, but not limited to: teamwork; professional communication; sharing knowledge and experience; aggregation and evaluation of information sources; presentation of the individual skills; discussions; construction of knowledge [4].

1.1 Context

If we assume that a good starting point for building an ESN are the best practices in the design and development of the popular social networking sites (like Facebook and Twitter) [2], then the features might include (in no particular order): groups; newsfeeds; calendars; blogs; link, file and photo sharing, polls; building a profile; building lists of friends/connections; personal and public discussions; engagement mechanisms;

content/behavior ratings; active intelligent environment – friend and content suggestions. Such an environment might have the following advantages:

- Increasing the engagement in the learning process – a set of tools and practices the users like and understand allows them to feel free to communicate and express their knowledge and skills. Appropriate game mechanics [3] might also provide a good balance between competition and cooperation.
- Building a network of professional connections – an ESN would allow finding and connecting to people that one might never interact with in other circumstances. The growth of the professional social networking [5] in the last few years shows the huge potential of this concept both for the individuals and the companies searching for the right employees.
- Putting the educational content in a "real world" context.
- Improving the connection and content discovery process.

1.2 Related Works and Possible Solutions

The initiative to try new educational approaches is usually left to individual professors and research teams [6]. The available paths from professor's point of view (assuming a willingness to initiate social learning) are: using the well-known environment of the popular social networking sites to communicate with the students (several studies at universities in different countries show that this approach seems to be ineffective for various reasons like: privacy issues, differences in expectations, and low participation rate of the lecturers [1, 7]); using services like Edmodo[1] or Ning[2], which allow the user to create a dedicated social network in the cloud; using a private ESN of their university (like University of Florida, Oregon State University and others [8]). The available paths (from student's point of view, assuming the professor didn't initiate social learning) are: creating or joining groups in SNS's, usually Facebook; studying on platforms like Coursera[3], Udacity[4]. These approaches might work in some situations but they cannot explore all of the possibilities a dedicated ESN would provide. That is why our approach has the following characteristics:

- Private ESN (from the university point of view);
- Integrated with internal e-learning systems and with external social networks, learning platforms and content repositories;
- Distributed – each instance of such an ESN is independent from the others and any central data storage [9];
- At the same time the communication with a component from the other instances is straightforward.

In an attempt to achieve the positive effects described above, the University of National and World Economy (UNWE), has made a decision to provide such an ESN for its professors and students. Since ELGG is the most popular open source software for SNS's,

[1] http://www.edmodo.com/

[2] http://www.ning.com/

[3] http://www.coursera.org/

[4] http://www.udacity.com/

already used for similar purposes at different universities, it fits the requirements put by UNWE and it has been installed, modified and extended for the purposes of the educational process in the university.

The goal of evaluating performance and identifying performance bottlenecks is to attempt to answer the following questions: can we use ELGG as a core, which can be extended to meet the requirements of a distributed, scalable, externally and internally integrated ESN; is the data model of this ESN scalable – can it handle growth in the user base, the content, and the relationships and can it handle complex operations on the data - UNWE is a large university (more than 500 professors and 16000 students) and if the network is used a lot, any performance issues might influence the user satisfaction with the system; to get to the full educational potential of an ESN it should involve a lot of participants from different universities and countries – can ELGG handle such activity. Since most of the popular SNS's do not use relational data models for performance reasons [10], the normalized relational data model of ELGG is the most likely bottleneck. The goal of the experiment described in this paper is to find out how the growth of users, relationships and content affects the performance of some of the most-used operations in such an environment. To check if the data model is a bottleneck, we are using statistics for CPU time and number of accessed rows (i.e. the effectiveness of the indexes) from the MySQL database engine.

2 Setup

The instance of ELGG is version 1.7.10 and the experiment was run on a server with 64-bit Windows and Apache and MySQL run as services, on a machine with a dual-core processor at 2.8GHz, 4GB of RAM. The system software and hardware details are relevant only to the concrete values – the conclusions are formed based on the ratio of the results of the different datasets. Since each operation is run multiple times with different starting user nodes, we could assume that any hardware side-effect would be reduced.

All of the test datasets consist of real students' and professors' data from the databases of UNWE. The "friendships" are generated based on: between one professor and another - mutual department or faculty, mutual subjects; between student-student – mutual majors, subjects, country and city of birth; between professor-student – mutual subjects. A group is generated for each subject taught by a registered professor and the students in that class are included in that group. Then the user participation is simulated by generating blog posts in the groups. By including more students and professors in the dataset it was possible to differentiate three scenarios:

Dataset A: 4168 registered users (539 professors and 3629 students), 14243 directed (not confirmed) friendships, 1146 groups, 55535 group memberships, 9875 blog posts.

Dataset B: 2084 registered users (235 professors and 1849 students), 3276 directed friendships, 1146 groups, 27848 memberships, 9875 blog posts.

Dataset C: 1042 registered users (111 professors and 931 students), 855 directed friendships, 1146 groups, 13875 memberships, 2565 blog posts.

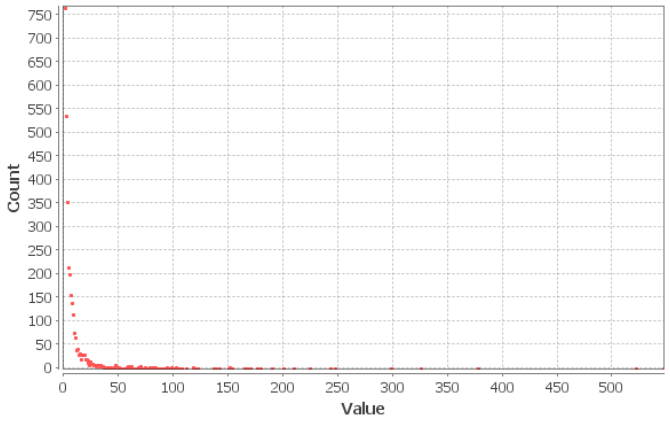


Fig. 1. The power-law degree distribution of dataset A. A few nodes have a high degree and all of the rest have a low degree. This distribution is typical for social networks.

To check if the generated social networks are realistic, it should be verified that the network is scale-free. The degree distributions have been calculated for each dataset (Fig. 1) – if the distribution is power-law we can assume that the dataset is appropriate for our tests, because this matches the distribution of the social networks in the working public SNS's [11].

3 Results

The operations (shown in Table 1) were selected by the expected frequency of their usage – these include the main menu elements of our implementation of ELGG and the most likely to be used educational components. Each operation was performed through the user interface of ELGG but the recorded CPU time only reflects the work done in the database. The results of the execution are presented in Table 1 and Table 2.

Table 1. Average CPU time (in seconds) per operation and dataset. Each operation is run at least 10 times with different starting user nodes.

Operation \ Dataset	A	B	C
adds a post to a group blog and previews it	0.5101	0.2733	0.1577
gets a list of her own friends	0.0709	0.1348	0.0595
gets the 10 most popular groups in the ESN	2.0686	0.7618	0.3341
gets the 10 most recent blog posts in the ESN	0.1303	0.1043	0.1686
searches for a person by name	0.1866	0.0988	0.0681
views a friend profile	0.0820	0.0519	0.0370
views own newsfeed	0.1468	0.0580	0.0317
views the friends of a friend	0.0565	0.0444	0.0357
views the front page of one of her own groups	0.6171	0.4283	0.1557
views a list of own groups	0.3268	0.0563	0.0328

In most operations the performance is decreasing linearly with the growth of the datasets – approx. 2 times growth in users, connections and content leads to approx. 2 times slower response from the database engine.

Another way to estimate the performance issues are the number of accessed rows in the database during an operation (Table 2). This shows how effective the indexes are. While in some cases the indexes are very effective, in most operations (esp. the complex ones, involving more objects) we are observing similar patterns to the CPU time values.

Table 2. Average number of accessed database rows per operation and dataset.

Operation \ Dataset	A	B	C
adds a post to a group blog and previews it	20656	15119	1775
gets a list of her own friends	1251	4490	1493
gets the 10 most recent blog posts in the ESN	27330	6970	4386
searches for a person by name	1658	1784	1128
views a friend profile	4070	1240	344
views own newsfeed	2526	1481	416
views the friends of a friend	2251	696	310
views front page of one of own groups	47885	21662	23732
views of a list of own groups	19852	1220	509

By analyzing the executed queries, the most important issue with the data model seems to be the metadata storage – it is organized to allow for easy development of extensions, which makes the querying of this data requiring a lot of table joins or separate select statements. While it is possible to get into further analysis of the causes behind these values, what becomes clear from this simple experiment is that the normalized relational data model of ELGG is a performance bottleneck and might cause significant issues with the implementation of the ideas presented in the first section of this paper.

4 Conclusions and Further Work

The normalized data model of ELGG works well for small social networks but might cause performance issues with the growth of users, relationships and content, especially considering complex data operations like analysis of the social graph (which are not available in ELGG). This makes it a weak candidate for the idea presented at the end of the first section of this paper. Since the bottleneck is the data model, the other available open source applications would probably have the same type of issues.

An ESN shouldn't "care" (from application logic point of view) how and where the data is stored, it should be able to work with independent abstractions and leave the storage mechanisms to what is available and appropriate in the current situation. The next step in this research should be to find out if it is feasible to build a data model, which is indifferent to the physical data storage and so it would overcome the performance limitations without making the application development too complex and too specific for a given environment. If some of the best practices from the developments in the area of enterprise social networking in the recent years are considered as well, such a model would also allow for learning workflow independence, for customizable integration patterns and for service type and location independence.

References

1. Wodzicki, K., Schwämmlein, E., Moskaliuk, J.: "'Actually, I Wanted to Learn': Study-related knowledge exchange on social networking sites," The Internet and Higher Education, vol. 15, no. 1, pp. 9-14, (2012)
2. Popescu, E., Cioiu, D.: "eMUSE - Integrating Web 2.0 Tools in a Social Learning Environment." (2011)
3. Vassileva, J.: "Toward Social Learning Environments," IEEE Transactions on Learning Technologies, vol. 1, no. 4, pp. 199-214, (2008)
4. Mason, R., Rennie, F.: E-learning and social networking handbook : resources for higher education. Routledge, (2008)
5. Bersin, J.: LinkedIn's Growth Continues: Fueling the Corporate Talent Machine, Forbes, http://www.forbes.com/sites/joshbersin/2012/05/07/linkedins-growth-continues-fueling-the-corporate-talent-machine/ (2012)
6. Butler, D., Sellbom, M.: Barriers to Adopting Technology for Teaching and Learning, Educause Quarterly, pp.22-28, (2002)
7. Grosseck, G., Bran, R., Laurentiu, T.: Dear teacher, what should I write on my wall? A case study on academic uses of Facebook, Procedia - Social and Behavioral Sciences, vol.15, pp. 1425–1430, (2011)
8. Organizations powered by Elgg, http://elgg.org/powering.php
9. Mitchell-Wong, J., Goh, S.K., Chhetri, M.B., Kowalczyk, R., Vo, B.Q.: Framework for Open, Distributed and Self-managed Social Platforms, IFIP, Vol. 283; Pervasive Collaborative Networks; Luis M. Camarinha-Matos, Willy Picard; (Boston: Springer), pp. 361–368, (2008)
10. Hoff, T.: Facebook Timeline: Brought to You by the Power of Denormalization, http://highscalability.com/blog/2012/1/23/facebook-timeline-brought-to-you-by-the-power-of-denormaliza.html (2012)
11. Adamic, L.A., Lukose, R., Puniyani, A., Huberman, B.: Search in Power-law Networks, arXiv:cs/0103016 [cs.NI], (2001)

"Cloud" business opportunities for mobile devices

Emil Denchev

Abstract. Integration among business information systems (ERP, CRM, SCM) of the manufacturer, distributor, retailer and customer is achieved by using mobile devices (Personal Digital Assistant (PDA), smartphone, etc.) The goal is to make the overall communication process among the participants in this market faster and more effective. In particular, these devices help the companies in better monitoring, analyzing and integrating data for the entire business cycle, starting from sales to payments in on-line mode. These devices improve the efficiency of the sales process and reduce costs significantly through optimization, including shorter delivery times, settling payments and reducing the unused storage area. The companies may also identify, attract and manage new clients and market segments using such devices. PDA can be used for - Business processing – taking and execution of client orders, Payment for the purchases, Clock and calendar for time planning, Address book for contacts, Internet access, viewing web pages, receiving and sending e-mails, Word processing, Video recording, listening to the radio, Global positioning system (GPS);

Keywords: Mobile devices, PDA, ERP, CRM, SCM, Mobile payment, NFC, SMS, WAP, USSD, Synchronization, Data Storage, VPN, SSL.

Introduction

Integration among business information systems (ERP, CRM, SCM) of the manufacturer, distributor, retailer and customer is achieved by using mobile devices (Personal Digital Assistant (PDA), smartphone, etc.) The goal is to make the overall communication process among the participants in this market faster and more effective. In particular, these devices help the companies in better monitoring, analyzing and integrating data for the entire business cycle, starting from sales to payments in on-line mode. These devices improve the efficiency of the sales process and reduce costs significantly through optimization, including shorter delivery times, settling payments and reducing the unused storage area. The companies may also identify, attract and manage new clients and market segments using such devices.

"Expect mobile payments to surge over the next few years, said analysts at research firm Gartner Inc. A new report from the company said the value of transactions employing mobile payment methods will top $171.5 billion this year. That's a growth rate of nearly 70 percent over 2011's $105.9 billion Gartner said. Additionally, users of mobile payments will grow from 160.5 million in 2011 to 212.2 million this year.

And Gartner said the advance of mobile payments worldwide will continue to be robust for the foreseeable future.

"We expect global mobile transaction volume and value to average 42 percent annual growth between 2011 and 2016, and we are forecasting a market worth $617 billion with 448 million users by 2016," said Sandy Shen, research director at Gartner" [1].

Mobile payments are based on the technologies made specifically for the purposes, viz. Near-Field Communication (NFC), Short Message Service (SMS), WAP or Unstructured Supplementary Service Data (USSD).

NFC uses a mobile device (PDA or Smartphone), equipped with a special chip which exchanges data with wireless reader, and records a deduction from the buyer's bank account. PDA is a small notebook without own keyboard, which is convenient due to its small size and PDA is also known as pocket computer or palmtop computer.

PDA can be used for:

- Business processing – taking and execution of client orders,
- Payment for the purchases,
- Clock and calendar for time planning,
- Address book for contacts,
- Internet access, viewing web pages, receiving and sending e-mails,
- Word processing,
- Video recording, listening to the radio,
- Global positioning system (GPS);

New PDA devices have color screens and audio capabilities thus making possible their use as mobile phones (smartphones) or portable players (media players). Many PDA devices have access the Internet, corporate networks (intranet) or external networks via Wi-Fi or Wireless Wide-Area Networks (WWANs).

Features

Standard PDA device has a touch-sensitive screen for input data, slot for memory card storage and integrated Wi-Fi and Bluetooth wireless technologies.

Operating systems for mobile devices, including smartphones

Opportunities for business use:

Data Synchronization

One of the important functions of the PDA is two-way data synchronization with enterprise software on a PC. This helps to maintain current information stored in your business software like CRM, Microsoft Outlook or ACT!. Data synchronization ensures that the PDA has the most recent list of orders, contacts, meetings and letters, allowing users to access the same information through the PDA, and/or a PC.

The synchronization prevents loss of information in case the device is lost or damaged. Another advantage is that data entry becomes much faster by using a computer while entering text using the touch-sensitive screen is considerably slower. So, transfer the data from PC to PDA can be made much more rapidly against manual entering all the data to the mobile device.

Synchronizing a PDA with a computer is performed with the synchronization software and HotSync Manager for devices with Palm OS and Microsoft Windows Mobile Center for Windows 7 PCs and devices with Windows Mobile.

These programs allow the PDA to be synchronized with a Personal Information Manager (PIM). This personal information manager may be an external program or an embedded program. E.g., BlackBerry PDA program comes with Desktop Manager, which can be synchronized with Microsoft CRM, Microsoft Outlook and ACT!.

There are also free tools as Dropbox (https://www.dropbox.com/) or SugarSync (https://www.sugarsync.com/) allowing the data synchronization. E.g., when using Dropbox each selected file is synchronized immediately between the "cloud" and applications on a desktop, laptop or any other mobile device used. Once the application is installed on a PC,

you can immediately make a backup copy of any selected file online. Then everything added to the folder sync is automatically copied in the "cloud".

Table 1.

	Operating systems	Company	Smartphone	Browser	Distribution
	Windows Mobile (Phone)	Microsoft	Samsung Omnia II, HTC Imagio	MS IE for mobile, Skyfire	Microsoft Store
	OS X Mobile	Apple	iPhone, iPad, iPod Touch	Safari	AppStore
	BlackBerry	Research In Motion (RIM)	All BlackBerry phones	Opera, Bolt	BlackBerry App World
	Android	Google	Motorola Droid, Google Nexus One, HTC Hero	Webkit	Android Market
	Symbian (Touch)	Nokia	Nokia, Samsung	Symbian Browser	Ovi Store
	Maemo	Nokia	Nokia Tablet	MicroB, Firefox Mobile	Maemo appstore
	Palm webOS	Hewlett-Packard	Palm Pre, Palm Pixi	Blazer	Palm App Store

Data Storage

PDA is used to store data accessible anytime and anywhere. Storage options are: locally on the PDA, and web based in the "cloud".

Web-based storage opportunities

Microsoft Office Live Workspace

Office Live Workspace for Microsoft Service allows to work with the known Office applications online likewise when documents are located on the local computer, including sharing Word, Excel and PowerPoint documents. It is not necessary to have a USB flash drive, as the service allows online storage of these three types of documents on the website Live Workspace (http://workspace.officelive.com). When a document is stored there, there is access to it from any device from any web browser. The content can be shared with other users in the workgroup and can be edited from any computer with Office installed package in the usual way, although the document is stored in the Microsoft servers - not on the local computer. To use Live Workspace service, it is necessary to remove Office Live Update for Office instrument from the Microsoft web.

Google Docs

Google Docs (http://docs.google.com) is an alternative to Google's package of the online office applications that includes several tools for synchronization between the usual desktop Office documents, applications and services on Google.

The most popular is Memeo Connect (http://www.memeo.com/memeo-connect-for-google-docs.php) which can synchronize any document from Mac or PC computer, including Word, Excel and PowerPoint documents with Google Docs. Here you will have the choice – either to edit the document locally with Office package or online with Google Docs. In both cases, Memeo Connect will support synchronization of changes made in Google Docs.

OffiSync

In a team work on a project with other managers from the remote locations, OffiSync (http://www.jivesoftware.com/about/companies/offisync) tool will offer better opportunities. OffiSync works almost the same way as Memeo Connect, while the only difference is that it allows more "cooperation" (collaboration) – so that it allows many users to work simultaneously on the same document with changes made by each user, visible to others working on it as well. For this purpose all users working on the project must have OffiSync installed.

Electronic calendar

Normally, the company employees use desktop electronic calendar helping them to organize meetings, important events and activities and generally maintaining order in their routine contacts and work. Taking this application to "the cloud" will allow access to the program always and anywhere, especially if it can be synchronized with a smartphone or other mobile device. Google Calendar (www.google.com/calendar/) and Google Calendar Sync plugin for Microsoft Outlook are used for this purpose. Once installed the calendar application, it takes care to maintain two-way synchronization between Outlook calendar and online wizard Google Calendar Wizard.

If you work with BlackBerry, iPhone, Nokia S60 smartphone or Windows Mobile, application Google Sync for your phone (http://www.google.com/mobile/sync/) enables wireless synchronization with your mobile device. So you can check the updates on the program and be up to date with any important meetings.

E-mail

The possibility of the permanent access to the office e-mail from any location at any time plays a key role. Many business users are already using the opportunities of “the cloud” services offered by Gmail.

By activating IMAP (Internet Message Access Protocol) support for Gmail account, it is possible to monitor everything through the office e-mail client. IMAP allows synchronization between a desktop e-mail client and Gmail inbox. To activate the service, you need to log in to Gmail account and follow the instructions to ensure relevant POP/IMAP page. This will provide access to the office e-mail from any mobile device that has a web browser.

Payment to business partners

Through the use of a small appliance for the mobile devices developed by "Square" company, a smartphone or a tablet can become a POS terminal to pay by credit card. So, any small business can accept credit card transactions as if he/she has the POS terminal.

At present the service is available only in the U.S. market, while over 2011 "Square" is reported over $4 billion in annual cash transactions through their devices [2]. This appliance works for iPhone, iPad and devices based on Google Android operating system.

Except for "Square", "PayPal" company offers mobile appliance PayPal Here. This is a small triangular credit card reader that plugs into the headphone jack. The service is almost identical to that offered by "Square", but has one major advantage - it can enter the world market much faster, as "PayPal" consumer base comprises more than 100 million people.

At present "PayPal" offers its mobile payments services only for iPhone, but application for Android is expected soon.

Development trends in mobile payments

"Square" company has developed technology for the mobile payments, which helps consumers pay for purchases without taking the smartphones out of their pockets. The only thing you have to do is to tell your name to the respective retailer and to make the purchase.

Due to this new service the users have do not need any appliance for their smartphones, but will need Card Case wallet version 1.1 application for the iPhone, which operates on the geolocation service base. If a user has installed the "Square" software on his phone and approached at least 100 meters away from the registered dealer, the respective seller automatically receives the alert on this. When this customer enters the store, the only thing he has to do is to tell his name.

The seller in his turn verifies the identity of the consumer with the photo from his iPad-application and gets the authorization to draw the money from the credit card registered with "Square" customer's account. The user must activate the Card Case wallet application and confirm the money transfer to the seller.

At present more than 800,000 people use the "Square" system for the mobile payments. The company services are available only in the U.S. Similar system iZettle is developed in Sweden.

Security

When smartphones and tablets users are connecting to Wi-Fi public wireless access zones there is a certain risk for intrusion and tapping.

"Best" practices are as follows:

1. To install personal firewalls (e.g. such as Windows Firewall) on all devices.
2. To use a virtual private network (VPN) access for the users to the company network. If the company does not have its own VPN, such external services as Hotsopt Shield or Witopia can be used. Embedded VPN clients can be installed in iOS (on the iPhone, iPad and iPod Touch) and Android devices.
3. To Use SSL encryption for external access to email (via mail client or the Web).
4. To install an antivirus software on the mobile device before adding any business applications.

5. To install only widely used programs tested and recommended by colleagues and friends.
6. To buy the applications only from the official "market" for the respective operating system. So, to avoid other programs out of control, Android deactivates parameters "Unknown sources" in the device settings.
7. While installing the application, the user sees a list of rights that the program must use for its complete functionality. It is recommended the application not to be installed if e.g. a game asking for the access to orders, contacts or permission for the data transfer.

References

1. http://www.mobilepaymentstoday.com/article/195057/Gartner-predicts-significant-mobile-payment-growth-through-2016
2. http://www.swift.com/resources/documents/SWIFT_white_paper_Mobile_Payments.pdf

MANAGEMENT INFORMATION SYSTEM IN BANKING SECTOR AS A PART OF CLOUD COMPUTING PLATFORMS

SILVIA TRIFONOVA

Department of Finance, Faculty of Finance and Accounting
University of National and World Economy - Sofia

ABSTRACT

The main objective of the paper is to analyze the management information system in the banking sector as a part of the cloud computing platforms. The tasks of the study are to define the cloud computing in different aspects, to describe the benefits of using cloud in core banking systems and to describe and analyze the groups of characteristics of software tools industry-leading companies. The central part of the paper is the analysis of the key benefits for the banking industry using cloud computing platforms and the major characteristics of these platforms in terms of applicability and security. The paper concludes with summarizing the results from the study.

INTRODUCTION

Cloud computing is one of the fastest-growing technologies in the past years. Business applications are the largest market for cloud services spending, with a gradual transition from on-premise to cloud-based services especially for general business applications. Today, there is a wide range of applications from general hardware clouds, application clouds and financial management to specific banking applications. The companies which planning to reap the benefits of cloud computing for competitive advantage must know and follow the development of software of software tools industry-leading companies. The objectives of this paper are the following:

- Defining cloud computing in different aspects;
- Describe the benefits of using cloud in core banking systems;
- Describe and analyze the groups of characteristics of software tools industry-leading companies.

1. DEFINING CLOUD COMPUTING

In the specialized literature the term "cloud computing" is not precisely defined. It is considered from various aspects. Several years ago, cloud computing is perceived as: “a collection of related concepts that people recognized, but didn’t really have a good descriptor for, a definition in search of a term, you could say” [8]

In terms of Software as Service (SaaS) Eric Schmidt from Google CEO department used “cloud computing” at a search engine conference in August 2006 to describe what they were doing: “the PaaS/IaaS connotations associated with the Google way of managing data centers and infrastructure”. Later in 2009 World Economic Forum started a research project about cloud computing, they had to grapple with the complexity and controversy of existing definitions, and ended up by adopting a broad definition of cloud computing that included “all kinds of remote services, from Software-as-a-Service to virtual machines” [6]

Gartner defines cloud computing as a "style of computing where massively scalable IT-related functions and information are provided as a service across the Internet, potentially to multiple external customers, where the consumers of the services need only care about what the service does for them, not how it is implemented. Cloud is not an architecture, a platform, a tool, an infrastructure, a Web site or a vendor. It is a style of computing" [4].

One of the most popular definitions, which gives a comprehensive review of cloud computing is on the University of California Berkeley RAD Lab (Reliable Adaptive Distributed Systems Laboratory): "Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services. The services themselves have long been referred to as Software as a Service (SaaS). The datacenter hardware and software is what we will call a Cloud." [1].

Broad definitions often focus on cloud computing like a part of the global network (World Wide Web) and organization's computing infrastructure: "In short, the cloud is the Real Internet, or what the Internet was really meant to be in the first place: an endless computer made up of networks of networks of computers. Even shorter: the Cloud is the Computer."[3] "Cloud computing is the distributed virtualization of an organization's computing infrastructure" [2]. Applications and files are hosted on a "cloud" consisting of thousands of computers and servers, all linked together and accessible via the Internet.

Technology Laboratory of the National Institute of Standards and Technology (NIST), tried to described the "standard definition" of cloud computing [7]: "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models."

1.1. Essential characteristics

On-demand self-service. A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service's provider.

Broad network access. Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, laptops, and PDAs).

Resource pooling. The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand.

Rapid elasticity. Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in.

Measured Service. Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported providing transparency for both the provider and consumer of the utilized service.

1.2. Service Models

Cloud Software as a Service (SaaS). The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (e.g., web-based email). The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.

Cloud Platform as a Service (PaaS). The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations.

Cloud Infrastructure as a Service (IaaS). The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (e.g., host firewalls).

1.3. Deployment Models:

Private cloud. The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.

Community cloud. The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third party and may exist on premise or off premise.

Public cloud. The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.

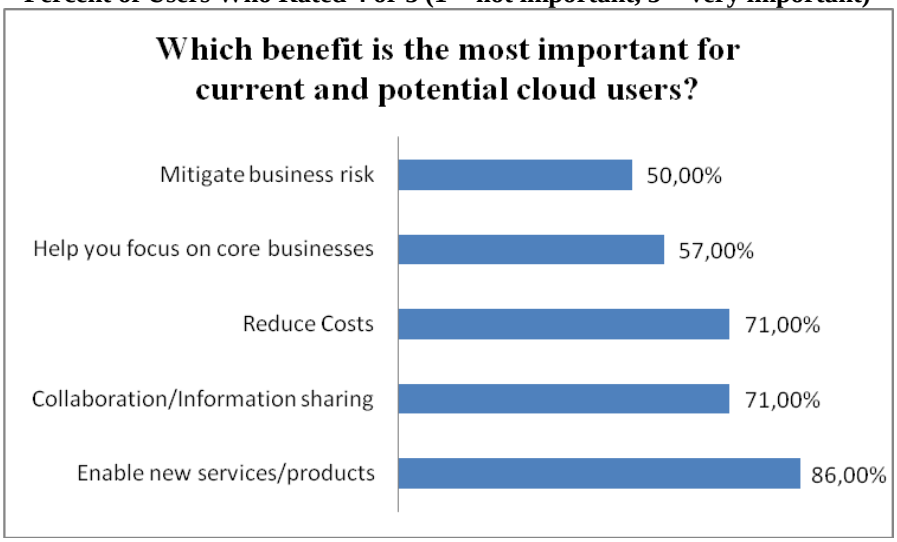
Hybrid cloud. The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-balancing between clouds).

2. MANAGEMENT INFORMATION SYSTEM IN BANKING SECTOR

The results of a research entitled "Exploring the Future of Cloud Computing"shows "Which benefit is the most important for current and potential cloud users". The research found that organizations using or planning to use cloud computing technologies do not view cost reduction as the most important potential benefit. While they are, indeed, lowering costs by accessing less expensive computer resources over the Internet, corporate and government

executives say the biggest advantage of cloud computing is better support for creating new products and services, and innovation in general (Figure 1) [9].

Figure 1: Which benefit is the most important for current and potential cloud users? Percent of Users Who Rated 4 or 5 (1 = not important, 5 = very important)

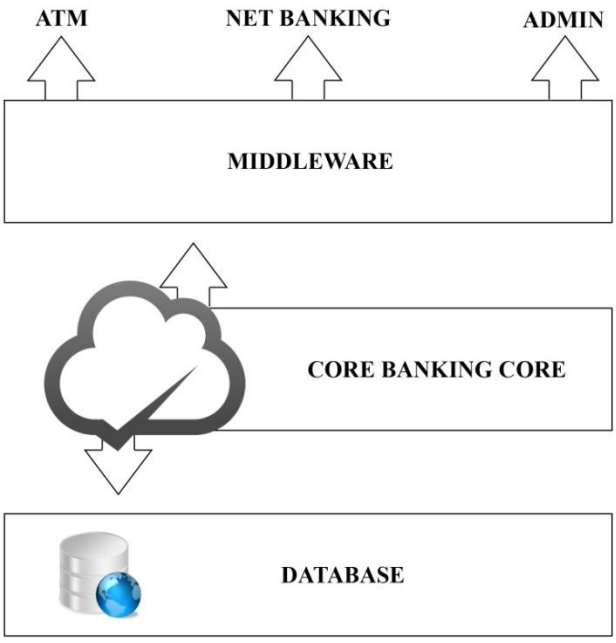


The Cloud Computing model of computing is a distributed application structure that segregates tasks or workloads between the suppliers of a service, called "Cloud servers", and service seekers, called "clients". Generally clients and servers communicate over a computer network on distinct hardware, but both may reside in one system. A server is a host machine that executes one or more server programs which share their resources with clients. A client machine never shares any of its resources, but requests for services of the server. Therefore clients start communication sessions with servers that are waiting for incoming requests.

For various banks and credit unions, constructing a technology infrastructure to satisfy current or evolving business needs is easier said than to implement. Budget, human resources, and/or capacity constraints influence the ability to constitute technology capabilities necessary to run and scale operations. Banks and credit unions have been leveraging service bureaus or outsourced core banking platforms for decades. In the present economic conditions, bankers are under acute pressure to cut costs and increase returns.

Figure 2 shows a "Banking Architecture using Cloud" which represents the banking scenario using cloud. As shown in the figure all the services like storage, applications, etc. are into the cloud. Thus providing the most accurate, powerful and adaptable business functionality, with a sophisticated and scalable architecture, that gives banks an unprecedented opportunity to reach its business goals [10].

Figure 2: Banking Architecture using Cloud



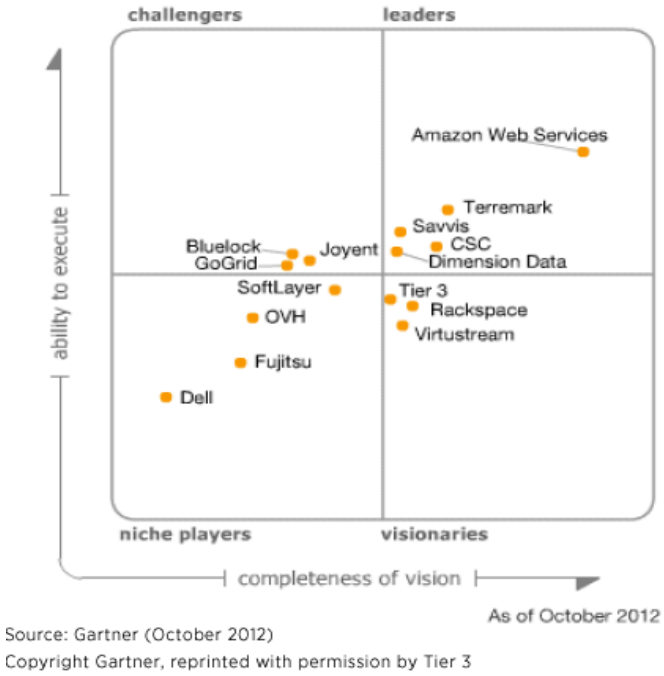
According to the aforementioned publication [10] the **benefits of using cloud in core banking systems are as follows:**

1. High costs incurred in running in-house data centers are removed.
2. Pay-Per-Use model offers operational agility when you need it.
3. Resources can be scaled efficiently and volume can be increased according to demand of the customer.
4. Banks can target on their core business as opposed to bothering themselves about infrastructure scalability.
5. Banks can be faster and more agile in generating new offerings. They need not worry about finding additional computing power.

3. CLOUD COMPUTING PLATFORMS – TECHNOLOGIES AND FUNCTIONALITIES

Companies offering cloud computing platforms are in a highly competitive environment. Confirmation of this is the matrix of Gartner data for 2012. As can be seen from Figure 3 in the quadrant "leaders" are referred to the following organizations: Amazon Web Services, Terremark, Savvis, CSC, Dimension Data.

Figure 3: Gartner's quadrant for Cloud Computing Companies



Source: Gartner (October 2012)

The market trend consists of an increasing use of cloud computing. Customers are mainly companies and institutions in the financial and banking sector and big commercial companies. The companies which planning to reap the benefits of cloud computing for competitive advantage must know and follow the development of software of software tools industry-leading companies (Table 1).

Table 1: Key characteristics of the Cloud Computing Platforms for the leading companies

COMPANY	CHARACTERISTICS
Amazon Web Services (AWS)	**Low cost** - no up-front expenses or long-term commitments; **Agility and Instant Elasticity** - massive global cloud infrastructure that allows you to quickly innovate, experiment and iterate: instantly deploy new applications, instantly scale up as your workload grows, and instantly scale down based on demand. Open and Flexible - **language and operating system agnostic platform.** Secure - **secure, durable technology platform with industry-recognized certifications and audits: PCI DSS Level 1, ISO 27001, FISMA Moderate, HIPAA, and SSAE 16** **Multiple layers of operational and physical security**
Terremark	**Power, Simplicity & Control** - Virtual machine goes from a template to a fully available computing resource; Prepare servers in advance; **Largest Number of Supported Operating Systems** - The Enterprise Cloud supports more than 450 operating system configurations including Windows Server 2003 and 2008, Red Hat Enterprise Linux, CentOS, FreeBSD, Solaris, SUSE Linux and Ubuntu. **Integrated Network and Firewall Management** - manage connecting your virtual machines to the Internet; the firewall management feature gives you the ability to modify and create firewall rule sets as needed for their environments; send firewall logs to a syslog server configured within your cloud environment, or externally if required **Integrated Load Balancing** - adding additional nodes to public IPs on the Enterprise Cloud. From the Web-based interface, customers can perform load balancing, network address translation (NAT), and port address translation (PAT), as well as combinations of those services **Private Network Integration** - integration with your own existing private network; configuration, address management, and security concerns associated with interconnecting this hybrid environmen; all network connections are made via secure channels (Utility LAN-LAN, Dedicated LAN-LAN, or Point-to-Point Circuit connections) that safely link only subscribed cloud resources to the enterprise network. **Multiple Environments in the World's Most Secure Facilities** - allowing you to create disaster recovery solutions and system redundancy across geographically diverse facilities. The Enterprise Cloud's Web-

	based console lets you manage multiple environments from a single interface. **Physical Device Integration -** the elasticity of the cloud combined with the security and compliance of dedicated hardware. **Security and Compliance** - a highly secure foundation **Copy Server -** ability to take a snapshot a server, creating a perfect block-level copy, or create master templates for later deployment of additional capacity.
Savvis	• **Software-as-a-service** • **Business continuity** • **Content management** • **Proximity hosting**
CSC	**Enterprise cloud infrastructure, infrastructure utilities and managed hosting** CSC CloudCompute Secure, resilient **cloud infrastructure service** designed to support production applications and workloads. CSC BizCloud™ **Private cloud** billed as a service, deployed from your data center or ours CSC BizCloud VPE Hosted private cloud with dedicated compute and networking and logically segregated storage delivered as a service. CSC Cloud DB Database delivered as a service for IBM DB2, MySQL or MS SQLincluding the software, infrastructure and run and maintain needed for your applications. CSC CloudDesktop™ for VMware View **Desktop as a service** on a private cloud delivered as a service CSC CloudIAM™ A complete identity and access management solution supporting both in-house and **Software as a Service (SaaS)**applications with single sign-on and centralized management. CSC CloudLab™ **On-demand Infrastructure as a Service (IaaS)** to meet the needs for temporary cloud compute and storage resources. CSC Managed Hosting **Enterprise-class hosting solutions** that supply the application infrastructures businesses need today and the architecture to support future goals. CSC Cloud IU™ for SAP A managed service delivering **SAP Basis support** and a cloud based infrastructure in a utility service by combining proven SAP expertise with our CloudCompute infrastructure. CSC Cloud IU for SharePoint Cloud Infrastructure and application run and maintain support on a consumption based delivery model for SAP or SharePoint.

	Security • Technology foundations, techniques, processes and staff to deliver the security and transparency required for cloud • Protected by defense-in-depth security framework that layers physical and logical security with access control and data protection • Built on the resilient Vblock™ platform from the VCE company that combines virtualization software from VMware, Cisco unified networking, security and computing from Cisco; and storage, security and management technologies from EMC • CSC leadership and compliance recognized by Vmware **Industry Leading Cybersecurity Practice** • Implementing mission critical security solutions • Innovative information security programs for compliance management, IT infrastructure and application protection, and disaster recovery • Sharing expertise and examining IT security as a technology strategy in the Digital Trust Report Series • Representatives on cloud and security • Committed to delivering cloud transparency to gain digital trust **Cloud Transparency and Digital Trust** • Sharing expertise and examining IT security as a technology strategy in the Digital Trust Report Series • Developed the CloudTrust Protocol to support the delivery of cloud transparency by cloud providers and develop digital trust in the cloud by IT consumers • CSC's Cloud Trust Protocol integrated by the Global Security Alliance as a pillar of their Governance, Risk and Compliance (GRC) stack
Dimension Data	**Managed Cloud Platform supports both public and private cloud services with a common user interface** • Cloud infrastructure (servers, storage, networking, virtualisation and operating system software) • CloudControl, a cloud management system that provides operational control and automation of cloud resource provisioning, orchestration, administration, and billing. **Flexibility is a cornerstone of our MCP architecture. Dimension Data offers several types of MCPs.** • Public MCP – Deployed around the world, Public MCPs are accessible by the general public and meet the requirements of a true public cloud; self-service, pay-per-use and with standardised services. • Private MCP – Deployed in a client data centre or a third-party data centre, a Private MCP is dedicated to a single client. Private MCPs deliver a fully-managed private cloud service. • Hosted Private MCP – Hosted by Dimension Data, the Hosted

	Private MCP is dedicated to a single client. Hosted Private MCPs deliver a fully managed private cloud service. • Provider MCP – Deployed in a service provider data centre or a third-party data centre, a Provider MCP is dedicated to a single service provider for the use of its clients or a single client. **API** REST-based API, you are able to: • Provision and de-provision Cloud Servers using Cloud Service Catalog or customer images. • Customise CPU, RAM, or storage quantities on each Cloud Server. • Import customer images. • Export images in Open Virtualisation Format (OVF). • Fully manage server capabilities that include: start, stop, delete, adding storage, or changing CPU/RAM. • Clone a server to create a virtual copy. • Provision VLANs, ACL-based firewalls, NAT, and load balancing to ease the building of multi-tier applications. • Administer multiple users, role-based permissions, reporting, and billing.

From the analysis made it can be drawn several groups of characteristics **in terms of applicability:**

- Fully available computing resource with largest number of supported operating systems;
- Customise CPU, RAM, or storage quantities;
- Clone a server to create a virtual copy;
- Public and private cloud services;
- Instantly deploy new applications;
- Integration with existing private network;
- Database delivered as a service;
- Desktop as a service;
- Physical Device Integration.

In terms of security can be drawn the following groups of characteristics:

- Multiple layers of operational and physical security;
- Physical and logical security with access control and data protection;
- Innovative information security programs for compliance management, IT infrastructure and application protection, and disaster recovery.

CONCLUSION

Choosing a cloud computing platform depends on the goals, objectives, data volume and other factors that banks identified as critical to their business. The success of any bank is in finding the optimal ratio - fast action, cost and risk weighted according to the strategies, goals and resources. The main benefits of using cloud computing are: 1) It reduces costs; 2) It optimizes the system resources 3) Faster and more agile in generating new offerings.

Cloud computing platforms are strategic potential that companies and banks should be evaluated as a means to achieve sustainable advantage.

REFERENCES

[1] Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R, Konwinski, A., Lee, F., Patterson, D., Rabkin, A., Stoica, I. & Zaharia, M. "Above the clouds: A Berkeley view of cloud computing."University of California Berkeley, Reliable Adaptive Distributed Systems Laboratory, Technical Report No. UCB/EECS-2009-28, (2009). [available at http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html]

[2] Cagle, K. "But what exactly 'is' cloud computing?" O'Reilly Broadcast, 2008. [available at http://broadcast.oreilly.com/2008/12/but-whatexactly-is-cloud-comp.html]

[3] Fingar, P. "Cloud computing set to unleash a perfect storm in business." Cordial Cloudburst, 2009. [available at http://www.cordys.com/ufc/file2/cordyscms_sites/download/6f5f4d1cfe8be9d78d972fa808d8702c/pu/cordial_fingar.pdf]

[4] Gartner Research. "Definition of Cloud Computing." In: Lori MacVittie, "Cloud Computing: It's the destination, not the journey that is important." DevCentral Weblog, 2008. [available at http://devcentral.f5.com/weblogs/macvittie/archive/2008/11/03/cloud-computing-its-the-destinationnot-the-journey-that-is.aspx]

[5] Mell, P. & Grance, T. The NIST Definition of Cloud Computing. U.S. Department of Commerce, National Institute of Standards and Technology, Information Technology Laboratory. Version 15, 10-7-09, 2009. [available at http://csrc.nist.gov/groups/SNS/cloudcomputing/cloud-def-v15.doc]

[6] Oram, A. "Cloud computing perspectives and questions at the World Economic Forum," WikiContent, 2009. [available at http://commons.oreilly.com/wiki/index.php?title=Cloud_computing_perspectives_and_questions&printable=yes]

[7] Urquhart, J. "Are the Feds the first to a common cloud definition? CNET News: The wisdom of clouds, May 10, 2009. [available at http://news.cnet.com/8301-19413_3-10237274-240.html]

[8] Willis, J. "Did Google's Eric Schmidt coin 'cloud computing'?" Virtualization Journal, 2009. [available at http://virtualization.sys-con.com/node/795054 and http://www.johnmwillis.com/cloud-computing/who-coined-the-phrase-cloud-computing/]

[9] World economic forum, Exploring the Future of Cloud Computing: Riding the Ne xt Wave of Tec hnology-Driven Transformation, 2010, [https://members.weforum.org/pdf/ip/ittc/Exploring-the-future-of-cloud-computing.pdf]

[10] Ramakrishnan,M., Ahamed Ali,S., Madhubala Kolluri,P.P, Auditing the Cloud Based Banking Systems for Ensuring Storage Security, 2012, [http://www.europeanjournalofscientificresearch.com/ISSUES/EJSR_79_2_15.pdf]

CLOUD COMPUTING STATISTICS OF INDIVIDUAL ALTERNATIVE TESTS WITH A SPHYGMOMANOMETER

Lubomir H. Lahtchev

Institute of Systems Engineering and Robotics, Bulgarian Academy of Sciences,
e-mail: lahchev@icsr.bas.bg , lahchev@iser.bas.bg

Abstract: The variety of non-invasive blood pressure measurement devices on different parts of the arm delineates questions relatively the accuracy of measurement, from one side, and the usefulness of their comparison, from another side. With an emphasis on these needs the current exploration is completed by measurements on a human individual with one and the same sphygmomanometer for upper arm. It suggests complementary experimental information with concern of the medical or general healthcare practice. The measured data for the blood pressure set abundance of analytic results of different trends within this topic. Hence, the statistic assessment of the BP and its signal, the data processing and analysis on them, are commonly fragmentary, intuitive, but united by the principles of the mathematic statistics and the theory of the probabilities.

Key words: statistics, sphygmomanometer, assessment, data analysis.

INTRODUCTION

The measurement of the blood pressure (BP) is a routine procedure at the medical practice. Variety of aerobic sphygmomanometers for non-invasive blood pressure measurement from the upper arm and the wrist is suggested on the market. On the one hand, the classic method of Korotkov exists at application on the upper arm only. Besides, several different opinions in [1] on BP measurement at the wrist characterize it with some discrepancy.

The International Protocol [1] gives also a single estimation: “Devices for self measurement that measure blood pressure at the finger are not recommended because vasoconstriction of the digital arteries can introduce substantial errors.”, which apparently does not characterize and refuse the finger sphygmomanometers at all.

It is already known that the International Protocol and the Standard EN 1060-4 [2] contain important disadvantages of a mathematical aspect, described in [3], which substantiate the question about their mathematical basis. As a result, many sphygmomanometers can be found accurate and other inaccurate without defined double signed ranges of existence, non-adequacy mathematic provisions on the early stage of data processing. In this paper, double-signed domains of existence of the deviations are introduced.

The current exploration is based on BP measurements, executed on an individual healthy human. The sphygmomanometer for the upper arm includes an aerobic cuff, which can also be located on the elbow and the forearm. Both the elbow and the forearm are locations for tests of the BP at alternative domains of the arm with a sphygmomanometer for the upper arm. Such tests are useful at both real measurement of the BP in alternative locations of the arm and exploration of the signal quality and device sensitivity. These can provide BP data for analysis extra, based on prepositions, logic expressions and having in

mind results of their statistic processing. As a general, the purpose of the paper consists of comparisons and inferences, based on statistically assessed parameters: systolic, diastolic BP and their corresponding statistic estimations.

CLOUD OF ASSUMPTIONS AND PRE-REQUISITES

The experimental measurements of the BP are completed with accordance to the recommendations of the producer [4] on a relaxed body of a human at the upper arms and also at two alternative points of the arms:

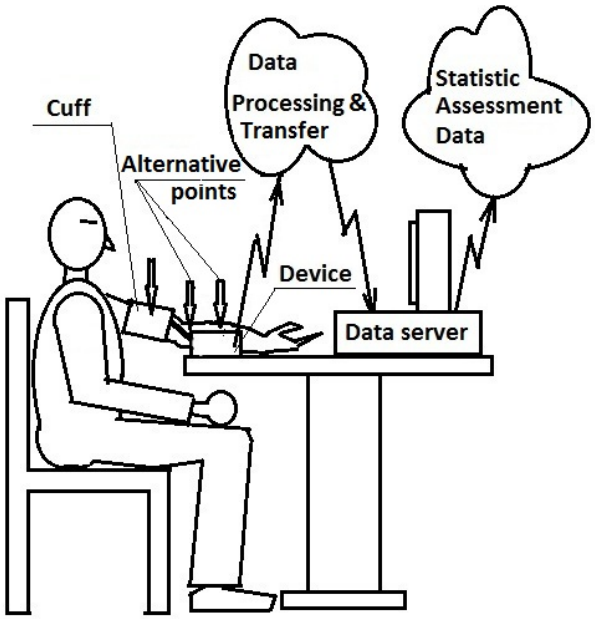


Figure 1. Scheme of BP-measurements with a sphygmomanometer for the upper arm at the recommended and alternative points, noted by arrows.

elbows and forearms. The scheme of BP-measurement is shown on fig. 1 and it is chosen in accordance to the principles of the singular micro- and macro-spaces: straight-linearity, parallelism, and balanced rotation. Portions of measurements consist of 6 entities with a kept five-minute interval between each one pair of them for one hand. The typified domains of measurements of the BP are ordered by consequence of their chronology as follows: left forearm, right upper arm, left upper arm, right forearm, left elbow, right elbow. For convenience, the corresponding arrays of systolic and diastolic BP have the next symbols: $p_{ylfSYSi}$, $p_{ylfDIAi}$, $p_{xruSYSi}$, $p_{xruDIAi}$, $p_{xluSYSi}$, $p_{xluDIAi}$, $p_{yrfSYSi}$, $p_{yrfDIAi}$, $p_{zleSYSi}$, $p_{zleDIAi}$, $p_{zreSYSi}$, $p_{zreDIAi}$ and the maximal index i is noted by N=81. Every one value within the array is mathematically noted as an integer α_i from the series n_{xi} – the upper arms, n_{yi} – the forearms, n_{zi} – the elbows, and the p(n) is the probability for $n_i = \alpha_i$.

It is correctly accepted the random background of the BP data and each one array represents issues of a statistically ordered by chronology data series.

The physic co-ordinate of the BP with a measure unit millimeter of mercury [mmHg] usually covers the distance of 40 – 200 mmHg, including 40 – 110 and some extra for the diastolic BP, and 60 – 180 mmHg for the systolic BP. Both sub distances of diastolic and systolic BP are separated on four defined closed zones and one open zone for each one of them, incorporated from the medical practice, for example [1].

As whole, twelve aggregate arrays by 81 issued values, called issues, for the left and right arms are classified by five zones of distribution. For the systolic BP they are 60-100, 101-130, 131-160, 161-180, more than 180 mmHg. For the diastolic BP they are 40-60, 61-80, 81-100, 101-110, more than 110 mmHg.

From the healthcare point of view, each one measurement of BP is unique and all issues have equal probability.

CLOUD OF THE HEALTHCARE STATISTIC ASSESSMENT ALGORITHM

For convenience, the mathematical basis of the estimation theory, presented in the monograph [5], is introduced here. The relative frequency $p_x(\alpha_i)$ of a unique data value about an issue $x=\alpha$ among an aggregate array can be computed as follows:

$$p_x(\alpha) = (1 / N) , N->\infty \quad (1)$$

where N – number of issues with $x=\alpha$ within the aggregate array. Computation of the eq. 1 suggests a probability for an issue, and corresponding aggregate array of values can be characterized by its matrix $\|p_x(\alpha_i)\|$=const. Thus the sum of probabilities remains equal to 1 [5]. With an account of the principle of the limit interrelation [5], the mathematical expectation of an aggregate array of the random value will be computed as a mean sum of addends, resulting from multiplication of the number of issue by its probability:

$$E\{x\} = \mu_x = \sum_{i=1}^{N} (\alpha_i \, . \, p_x(\alpha_i)) = \sum_{i=1}^{N} (\alpha_i \, . \, 1/N) \quad (2)$$

where 1/N can be drawn outside the sign of sum.

Having in mind this first statistic estimation of the aggregate array, the corresponding matrix of statistic standard deviation of first order can be computed:

$$\left\| E_i\{(x - \mu_x)\} \right\| = \left\| (\alpha_i - \mu_x) \right\| \quad (3)$$

where i=1, 2, 3 … N.

Certainly, the mean sum of differences:

$$E\{x - \mu_x\} = (1/N) \, . \sum_{i=1}^{N} (\alpha_i - \mu_x) \quad (4)$$

will define this second statistic standard estimation of the aggregate array. Hence, the second statistic estimation serves for computation of the dispersion of the random value:

$$E\{(x - \mu_x)^n\} = \sum^{N} ((x - \mu_x)^n \, . \, p_x(\alpha_i) = \sum^{N} ((x - \mu_x)^n \, . \, 1/N) \quad (5)$$

i=1 i=1

where n=2 for the square order of this statistic standard deviation and 1/N can be drawn outside the sign of sum.

The mean squared statistic standard deviation, expressed as follows:

$$\sigma\{x) = (E\{(x - \mu_x)^n\})^{1/n} \quad (6)$$

for n=2, provides third statistic estimation of the random value of an aggregate array, defined by dimension unit, which often points the main integral characteristics of it.

For the purpose of the statistic and educative analysis, an introduction of a set of differential thresholds about the first order deviations – random small numbers is recommended. Here, the values of differential thresholds are incorporated from the International Protocol [1], but the mathematical domains of existence are of a better double-signed accuracy, as follows:

$+5 \geq +G_{11} \geq 0$; $0 > -G_{12} \geq -5$; $+10 \geq +G_{21} > +5$; $-5 > -G_{22} \geq -10$; $+15 \geq +G_{31} > 10$; $-10 > -G_{32} \geq -15$; $+G_{41} \geq +15$; $-15 > -G_{42}$.

where +, – are the algebraic signs of the deviations of issues by separated double-signed ranges of their existence.

The estimation status of the BP random value is defined by ratios of index numbers of the issues, ceasing to a defined range of deviation, to their whole number as follows:

$$\bar{x} = (J_x / N) \cdot 100\ [\%] \quad (7)$$

where J_x – maximal number of issues x, which cease within a zone G, N – whole number of the issues in the aggregate array.

EXPERIMENTAL RESULTS BY DEVIATION DISTRIBUTION OF ALTERNATIVE TESTS ON THE ARM

The available data about twelve aggregate arrays were statistically ordered and sorted in accordance to algorithm of the corresponding computing cloud. The final results about average mathematically expected values and the corresponding mean squared statistic standard deviations are shown on table 1. They are foundation about general inferences:

GI1. The BP-values within the forearms exceed the BP-values within the upper arms and the lasts exceed the BP-values within the elbows.

GI2. The squared statistic standard deviations are minimal for the BP measurements at the upper arms, and with accordance to the results.

GI3. The squared statistic standard deviations from computations on the data about systolic BP at the forearms are maximal among the experiments and exceed the maximal step of 3 mmHg of the sphygmomanometer, but they could be results of a higher sensitivity to the oscillations of the systolic BP.

Table 1. The average values – mathematical expectations about BP and corresponding mean squared standard deviations.

		right arm	MSSD	left arm	MSSD
upper arm	systolic BP	122.22	±1,336	119,43	±1,298
	diastolic BP	79.21	±1,1865	76,89	±1,175
Forearm	systolic BP	127,86	±1,7092	122,69	±1,5318
	diastolic BP	85,97	±1,321	82,037	±1,3049
Elbow	systolic BP	112,77	±1,2417	110,87	±1,3679
	diastolic BP	77,037	±1,3765	76,07	±1,4867

GI4. The squared statistic standard deviations from computations on the data about diastolic BP at the elbows are maximal among the experimental results of diastolic BP at other locations and they exceed the corresponding squared standard deviation about systolic BP at elbow only, but these circumstances together with the decreased averages and distance between average systolic and diastolic BP data could be results of a low sensitivity to both systolic and diastolic BP.

GI5. The highest averages of the BP measurements are computed for the locations at the forearms about both systolic and diastolic BP, which together with GI3 are a basis for explanation the high sensitivity of the Korotkov's method to the increased readings and or oscillations at those locations of the aerobic cuff.

GI6. The equal data results together with the minimal mean squared standard deviations of the upper arms (GI3) at equal sensitivity of both systolic and diastolic BP are a foundation for analysis of the BP at the medical healthcare diagnostics.

The distribution of deviations of the systolic and diastolic BP by percentages of issues, ceasing to the separated ranges along with the locations of the cuff are numerically shown on the table 2 and illustrated on figures 2, 3, 4.

Table 2. Deviation distribution of the systolic and diastolic BP by the sixth locations of the cuff.

Locations of the cuff at measurement of the Systolic/Diastolic BP	Percentage of the issues within separated ranges 0±5mmHg	Percentage of the issues within separated ranges ±5±10mmHg	Percentage of the issues within separated ranges ±10±15mmHg	Percentage of the issues within separated ranges outside the ±15mmHg
Systolic BP of the right upper arm	41,9753086	32,0987654	22,2222222	3,7037037
Diastolic BP of the right upper arm	48,1481481	33,3333333	12,345679	6,17283951
Systolic BP of the left upper arm	48,1481481	34,5679012	13,5802469	3,7037037
Diastolic BP of the left upper arm	55,5555556	25,9259259	14,8148148	3,7037037
Systolic BP of the right forearm	37.037037	28.3950617	24.691358	9.87654321
Diastolic BP of the right forearm	44.4444444	28.3950617	18.5185185	8.64197531
Systolic BP of the left forearm	39.5061728	28.3950617	18.5185185	13.5802469
Diastolic BP of the left forearm	44.4444444	33.3333333	17.2839506	4.9382716
Systolic BP of the right elbow	43.2098765	37.037037	14.8148148	4.9382716
Diastolic BP of the right elbow	43.2098765	25.9259259	20.9876543	9.87654321
Systolic BP of the left elbow	41.9753086	28.3950617	16.0493827	13.5802469
Diastolic BP of the left elbow	37.037037	25.9259259	29.6296296	7.40740741

The graphics on fig. 2 demonstrate in detail the steep slope straight-linear distribution of deviations of a high triangle with a long hypotenuse.

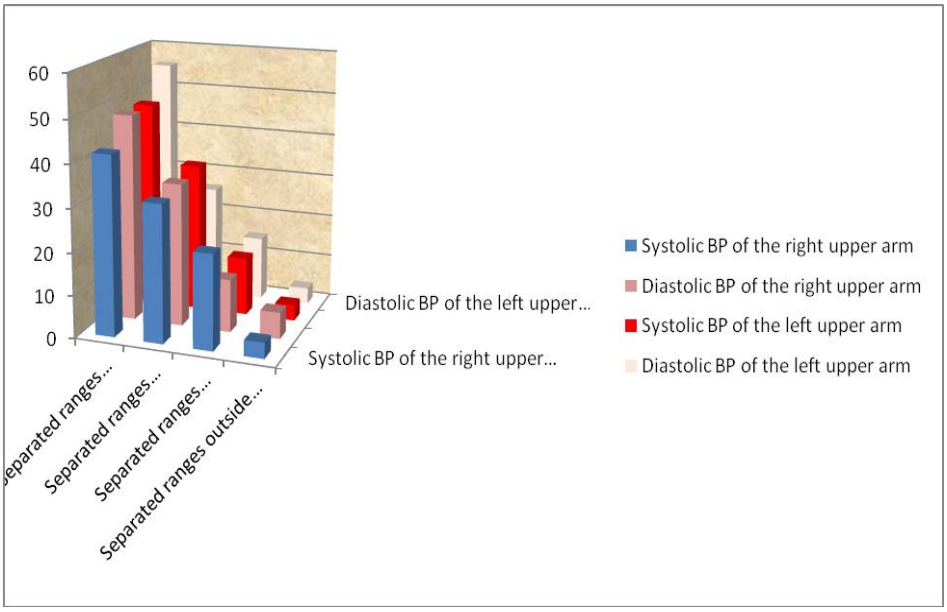


Figure 2. Deviation distribution of the systolic and diastolic BP of the upper arms by percentages of the issues within the separated ranges of deviations.

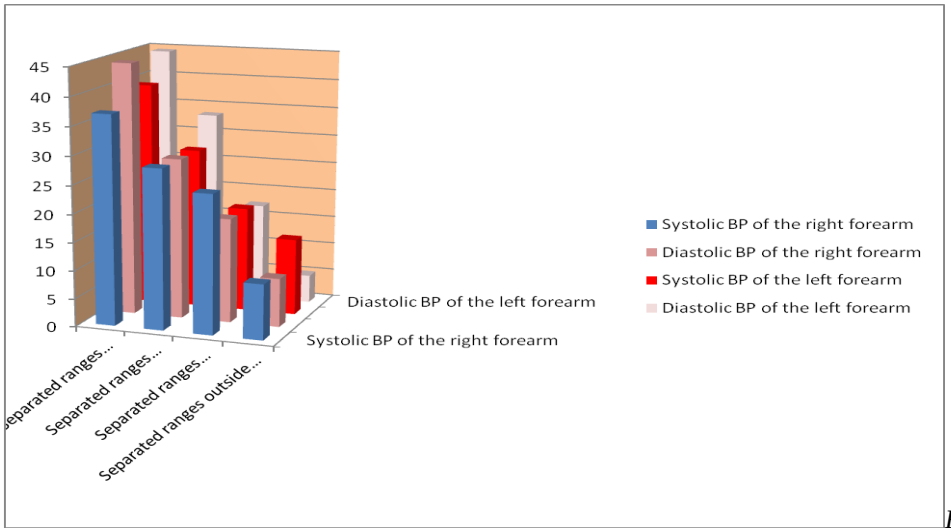


Figure 3. Deviation distribution of the systolic and diastolic BP of the forearms by percentages of the issues within the separated ranges of deviations.

The triangular form of distribution repeats within small angular deviations at readings among all four aggregate arrays of the BP at upper arms. This circumstance confirms the stability of the signal about BP at the upper arm.

The distribution of deviations for the forearms has increased dispersion toward higher values of deviations, seen on fig. 3, but still serving proportional triangular distribution.

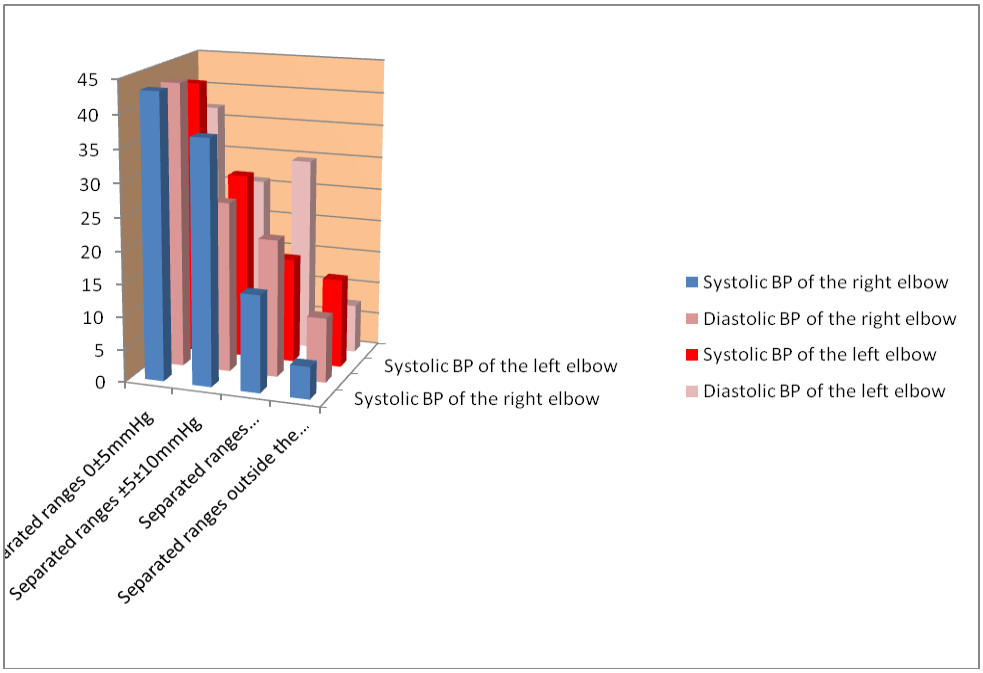


Figure 4. Deviation distribution of the systolic and diastolic BP of the elbows by percentages of the separated ranges of deviations.

Some wave forms of the distribution of BP-deviations at the elbow and their different varieties, which increase within the ranges of large deviations, shown on fig. 4, point the disadvantages of this location for non-invasive measurement with sphygmomanometer. Increased dispersion of the readings at the elbows substantiates the increased dispersion of the BP-signal at this location.

CONCLUSION

The paper introduces double-signed thresholds about the deviations of the BP-data, obtained from non-invasive sphygmomanometer. The BP-data are statistically processed with use of a statistic algorithm with equal probability of the issues, which characterize the healthcare point of view on the human individual. For this aim general distribution of data and distributions of deviations of the BP-data by locations of the aerobic cuff of a sphygmomanometer for upper arm had been computed. The alternative tests of non-invasive measurement of the BP on an individual aid the understanding of sources of non-sensitivity at it, what is numerically and graphically assessed and demonstrated. The inferences are also a base for a hypothesis, which is scientifically substantiated and estimated by likely-hood-based similarity in [6]. It supports the idea, that BP in the forearm exceeds the BP in the upper arm with a probability of 0.7222.

The cloud-computing statistic algorithms are valued on Excel.

ACKNOWLEDGEMENT

The paper is a continuation of the completed project MI 1509/2005, supported by the Bulgarian national fund of scientific explorations. It also contains results on estimation of integrated discrete-continuous complex processes, represented by the above analysis.

REFERENCES

1. E. O'Brien; Th. Pickering; R. Asmar; M. Myers; G. Parati; J. Staessen; Th. Mengden; Y. Imai; B. Waeber; P. Palatini; W. Gerin; N. Atkins. *International Protocol for Validation of Blood Pressure Measuring Devices in Adults.* Blood Pressure Monitor., February 2002, vol.7, issue 1, pp. 3-17.
2. British Standard BS EN 1060-4: 2004. *Non-invasive Sphygmomanometers – Part 4: Test Procedures to Determine Overall Accuracy of Automated Non-invasive Sphygmomanometers.* 36 p.
3. L. Lahtchev. Cloud Computing Individual Assessment of a Medical Device for the Education. Proc. of the Intern. Conf. on Appl. of Inf. and Com. Tech., December – 2011, UNWE, Sofia, pp. 503-512.
4. Tensoval® Duo Control. *Instructions for use*. Published on the site: www.hartmann.info , www.tensoval.de 2008, pp. 126 – 147.
5. Э. Сейдж, Дж. Мелс. *Теория оценивания и ее применение в связи и управлении.* Москва, "Связь", 1976, 496 с.
6. L. Lahtchev. *Likely-hood-based Educative Data Analysis of Alternative Tests with a Sphygmomanometer*. Proc. of Intern. Conf. on Appl. of Inf. and Com. Tech. and Statist. in Ec. and Educ., October – 2012, UNWE, Sofia.

Cloud Computing in Enterprise Content Management

Veselin Popov[1], Angelin Lalev[1]

[1] D. A. Tsenov Academy of Economics – Svishtov, Bulgaria

Abstract. Contemporary social and business environment expects and needs from technology and management solutions to deal with the processing of exponentially growing content. At the current stage Enterprise Content Management provide an effective strategy for linking people, processes and information in an organization. This strategy, however, requires strong technology support. Such support we see in cloud computing.

Cloud computing sets a new stage in the revolutionary development of information technology. Tendencies in the development and management of organizational technology show that very soon the majority of computer calculations will be carried out not in the office but in the "cloud." Increasingly, Web-based email services, social networking sites, photo and video services, online applications are distributed through cloud computing and are available to users through browsers, smartphones and more devices.

Putting ECM in the cloud enables organizations to collect, correlate and analyse data by disparate ways now. Cloud computing offers tremendous capacity to use technology for connecting disparate people, for storing and sharing information.

The purpose of this report is to examine the new opportunities provided by cloud computing to enterprise content management (ECM) as well as potential problems to adapt theology.

Keywords: Enterprise Content Management, Cloud computing, Security

Introduction

ECM is "a set of technologies, tools and methods used to collect, manage, store, preserve and deliver content across an enterprise"[5]. It provides a modern integrated perspective on information management within the organization. ECM helps the organizations to manage people, processes and content in a logically consistent manner. The application of Web 2.0 technologies for this purpose is known as ECM 2.0.

Every day, without even realizing, users spend considerable time in the cloud, using Web-based email services, social networking sites, photo and video services and online applications – all of these are services that are distributed over the components of cloud computing (CC), and can be accessed through browsers, smartphones and other consumer devices. In the near future it is expected that the majority of computer activities that are now performed on PCs at home or at work will take place in "the cloud," including those related to content management.

[5] AIIM - Association for Information and Image Management. The Global Community of Information Professionals. http://www.aiim.org/

This report attempts to outline the advantages of CC provides the plant with respect to the ECM, and possible barriers to its use.

The term "cloud computing" is a metaphor representing an abstraction of the underlying structure of the Internet, traditionally represented by a cloud (Fig. 1).

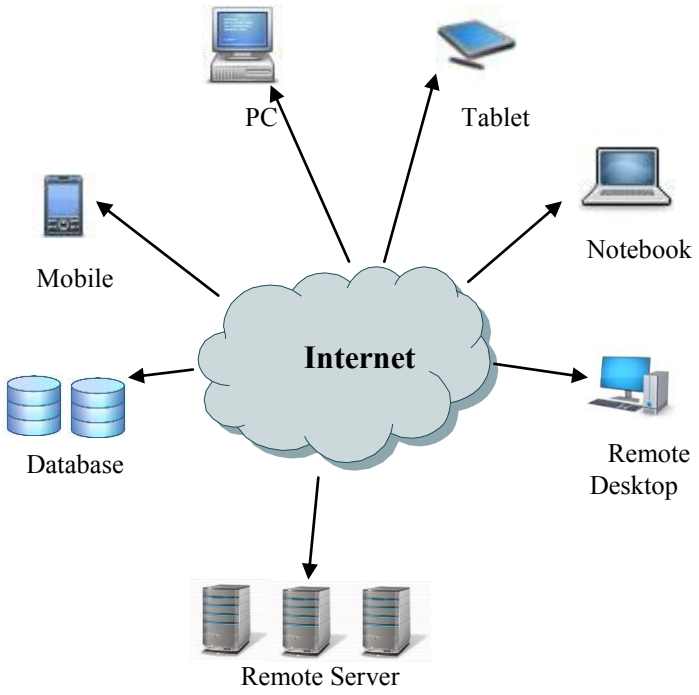


Fig. 1 Cloud computing

CC technology offers a variety of services for information storage and share, database management and deployment of Web services. The main idea behind the development of this technology is **cost efficiency** with the ability to engage computer services provided by a third party (supplier) rather than to build and maintain their own physical infrastructure. Comparing the benefits of hiring a technological infrastructure with the capital expenditure required for its purchase, the organizations more frequently orientate themselves towards SS for content management and services to ensure cooperation. The assignment of tasks such as document management, archiving, workflow, etc. to third-party enables organizations to use their time and resources for activities that create new value for businesses.

In addition to changing the way of acquiring technology, the concept of cloud computing changes the way of their ***consumption*** as well, as consumers are not required to have experience and knowledge in the management of IT infrastructure. Cloud computing provides highly scalable and available computing resources and simultaneously minimize the potential interaction of the user with technology.

ECM in the Cloud

Putting ECM in the cloud helps the organizations reduce infrastructure investments and administration costs. At the same time, they provide many new opportunities for the organization and management of the business.

Content in the cloud. With the advent of cloud computing the calculations were involved in ECM to effectively and efficiently manage the growing avalanche of information in digital format and take advantage of economies of scale as well. Cloud platforms give the opportunity to meet the specific hardware requirements of the ECM in the most effective manner. As a result, the organizations can provide themselves with new applications without having to purchase hardware, software, and consultants who need to create and manage enterprise systems. Outsourcing of tasks like for instance managing email or web content is a flexible solution. Simultaneously, we must note the problem that the access to data hosted in the cloud and transmitted from remote locations can slow the connection.

Collaborate in the cloud. Social computing applications (wikis, blogs and social networks) are increasingly perceived by business as production tools that provide a flexible and attractive way of communication and collaboration between users. By accessing the web from a desktop computer, laptop or mobile device, users collaborate in the cloud quickly, safely and easily. "The revolutionary for cooperation in the cloud is that it crosses geographic, cultural and organizational boundaries, offering a solution that sets the user in control." At the same time in this respect there are some valid reasons to concern about the data security, their transparency and compliance with the regulatory standards on user-generated content, that appears to be the product of collaboration in the cloud.

Process in the cloud. Modern cloud platforms enable organizations to create mashups, which combine the content or functionality from external sources and create a new service that is targeted at a specific business problem. Working in the cloud, the organizations use these services to connect with the operations of other companies that also use the cloud. This allows them to define business processes and to include other businesses from the cloud on their own processes.

Security of cloud-based ECM systems

ECM systems by design contain very sensitive information – the internal documentation of the organization, whose leakage may lead to direct and indirect financial losses. This fact makes the achievement of adequate level of information security within the bounds of the ECM system highest priority of the organization. This priority gets in conflict with the outsourced model of cloud computing. The latter creates inherent security risk by storing the documents of the organization outside the boundaries of organization. In many cases this outsourcing causes the data being stored across the national borders and even – across the borders of large political and economic unions like EU.

The question how to adequately protect the data of ECM systems in such environment comes quite naturally. So far though, there is no universal solution to the

problem, but there are many ideas how to partially mitigate the risks, associated with ECMs in the Cloud.

One possible measure in this direction, used often in many real-world cloud usage scenarios, is to take care not to put the most sensitive company documents (such as business plans, innovative engineering plans and designs, customer lists etc.) in the Cloud. This seems easy enough when the amount of the documents is small, but the typical usage of ECM usually implies cases where the amount of documents actually so big, that it presents serious logistical challenge. In this situation is very easy to upload by mistake sensitive documents into the Cloud. To avoid such situations, is often useful to design and use classification of the documents, based on its sensitivity, and use it to systematically label the company documents. Such approaches are described in many sources [5,6,7,8], but none is explicitly related to an ECM system.

Other possible measure concerns the use of traditional symmetric encryption to encrypt the contents of the files, uploaded into the ECM system. This approach actually has potential specifically in relation to ECM systems, since quite many of the documents in an ECM system today represent unstructured information which has not or doesn't need to be represented in textual form. (i.e. images, of documents, stored and retrieved much more faster than their paper counterparts). Since the search on unstructured information will probably be limited to the metadata of the documents, the encryption of the actual document won't influence the search significantly.

A relatively new idea, combining both of the previous approaches is the so called "Information Rights Management" (IRM)[6]. IRM combines encryption and data labeling to achieve more robust system of protecting enterprise data. In a scheme, technically similar to DRM, the contents of the document is stored in encrypted form, together with its usage rights. The software of the organization is then configured to decrypt the content of the document, but only if the user has the necessary access and usage privileges. In such way, every document, uploaded into the Cloud can be easily made inaccessible to potential malicious cloud provider[7]. Of course this has the same limitations, in relation to the search as the ones described above. If the cloud provider cannot process the contents of the document, he cannot do any searches on behalf of the user in the contents of the stored documents. This limits the usefulness of this measure to documents, whose contents won't be subject of searches anyway.

The two measures represent an approach which is quite specific and related to some serious compromises. But at the present moment, there is little else that can be implemented in practice. This situation may change in the several few years thanks to the appearance of new unconventional encryption mechanisms which realize the so called "fully homomorphic encryption". Such encryption scheme would allow creation of programs that run directly on the encrypted form of the data. This in turn would allow cloud providers to process any data without having to decrypt it. Ultimately, this would lead to full and universal protection of the information, stored in the Cloud, including the data of ECM systems.

Up to 2009, the progress of development of fully homomorphic encryption has been slow. For more than 30 years the very question if such encryption schemes were at all possible was unsolved. The appearance of the first such scheme [9], followed by several others [10,11], changed this and created very dynamic and quickly moving area of research.

[6] Sometimes called "Enterprise Digital Rights Management" (E-DRM).

[7] Similar in principle idea was proposed from the community, associated with developments of GRID and Grid Security Infrastructure (GSI) several years ago.

The problem with all proposed schemes so far is that they introduce too much overhead in the processing of data and make cloud computing infeasible from economic standpoint at least. Some of those schemes were improved multiple times, which led them closer to achieving economic efficiency. Still a serious breakthrough is needed to achieve this goal.

Conclusions

Cloud computing offers new possibilities for connecting positioned dispersed users and a huge capacity for storing and sharing information. Putting ECM 2.0 in the cloud, i.e. the use of ECM as a service, enables organizations to collect, correlate and analyse data in a new way and to use the computing capacity of a new dimension.

At present the cloud services involve a threat of exposure of confidential information and it is necessary that the organization, which will use them determines exactly what is expected from the implementation of cloud information management (information governance) and IT strategies for compliance (compliance strategies).

For now, the rate of adaptation of cloud computing by organizations is hampered by the need to ensure security, privacy and reliability, and key technologies such as SaaS and more are still under development. However, it is expected that in the coming years, organizations will develop their own capacity or will choose hybrid cloud solutions.

References

1. Hamid R Motahari-Nezhad, Bryan Stephenson, Sharad Singhal. Outsourcing Business to Cloud Computing Services: Opportunities and Challenges, HP Laboratories, HPL-2009-23.

2. Jenkins, Tom. Managing Content in the Cloud, October 2011, Canada, Second Edition, ISBN 978-0-9730662-8-9

3. Cameron, S. Enterprise Content Management - A Business and Technical Guide, Charlesworth Press, 2011, ISBN 978-1-906124-67-0

4. AIIM - Association for Information and Image Management. The Global Community of Information Professionals. http://www.aiim.org/

5. Bell, David and LaPadula, Leonard. Secure Computer Systems: Mathematical Foundations. MITRE Corporation.

6. Common Criteria for Information Technology Security Evaluation (ISO/IEC 15408), http://www.commoncriteriaportal.org/cc/

7. Standard Security Label for Information Transfer (FIPS PUB 188), http://csrc.nist.gov/publications/fips/fips188/fips188.pdf

8. Oracle Label Security http://www.oracle.com/technetwork/database/security/ds-security-label-security-11gr2-134491.pdf

9. Gentry, Craig. A fully homomorphic encryption scheme, PhD dissertation, Stanford University, 2009 http://crypto.stanford.edu/craig

10. Lyubashevsky, V., Peikert, C., Regev, O. On Ideal Lattices and Learning with Errors over Rings. EUROCRYPT 2010. http://www.iacr.org/cryptodb/data/paper.php?pubkey=24000

11. Dijk, M., Gentry, C., Halevi, S., Vaikuntanathan, V. Fully Homomorphic Encryption over the Integers. EUROCRYPT 2010. http://eprint.iacr.org/2009/616

MULTITENANCY IN CLOUD COMPUTING

Vanya Lazarova
vlazarova@unwe.bg, UNWE, department of IKT

Abstract. The report will be considered on an essential feature of cloud computing - Multitenancy - the ability to share resources among a large community of users. Multitenancy occurs specifically in the three Delivery Models of cloud computing. Only multitenancy in Software as a Service model will be discussed the report. Multitenancy occurs specifically in the three Cloud Deployment Models of cloud computing. I will discuss its implementation in the public and private clouds.

Keywords: cloud computing, multitenancy.

Introduction

Since past 3-4 years Internet is not the same place as it was before. Many technologies are revealed combined with name semantic web or web 3.0.

Cloud computing is a new technology not based on the semantic web. Cloud computing technology rather transfers interactions in the real economy to the internet – such dealings as hire, provider, vendor, services, fit more to the real world.

Cloud computing is an outcome of the transformation the www from a tool of communication to a tool of data processing.

Cloud computing is a sharing a software, resources, and information basely by Internet and basely by access on demand.

This report examines one of the main characteristics of cloud computing - Multitenancy - the ability to share resources among a large community of tenants.

"Cloud" is a versatile, freely defined structure that integrates all the means available and delivered via the Internet - hardware and software systems of remote data centres that provide those services.

Cloud computing have changed the way information systems were regarded - they split the data processing, data storage and presentation of data from their original place of location. Services effected through this paradigm become increasingly popular.

Cloud computing offer many different options. They allow network access on request, as well as a common set of configurable computing resources. The operation can be performed rapidly with minimal management effort or interaction with the service provider. Of particular benefit is the flexible infrastructure platform providing cloud computing and the ability to change the computing resources as needed. Tenants are offered a model of "pay as much as you consume" of the resource they need. Purchasing services from a cloud may enable business decision makers to save money and allow companies to focus on their core business - enticing offers in the current economic climate.

Multitenance means the option of tenancy of the same services among a large community of tenants for various purposes.

However there are a lot of risks for its use by many tenants. These consist mainly of insufficient segregation of data in databases, insufficient level of identity management and access to data and inadequate recording and monitoring of the activities.

Multitenancy in respect to the model "software as a service"

Multitenancy defines itself specifically in the three services delivery models:

Software as a Service (SaaS) - delivering of applications to tenants.

Platform as a Service (PaaS) - delivery of platform to the tenants (usually at organizational and companies' level).

Infrastructure as a Service (IaaS) - delivery of resources to tenants.

The report demonstrates the effects of Multitenancy in model SaaS.

Information technology "cloud" and moving applications to the cloud created the need for different approaches in the processes of storage and maintenance of the data related to client's application. Here are the options available:

1.Single Tenant which uses its own database - in this approach the tenants are offered separate databases but it is not allowing scalability and flexibility of the application.

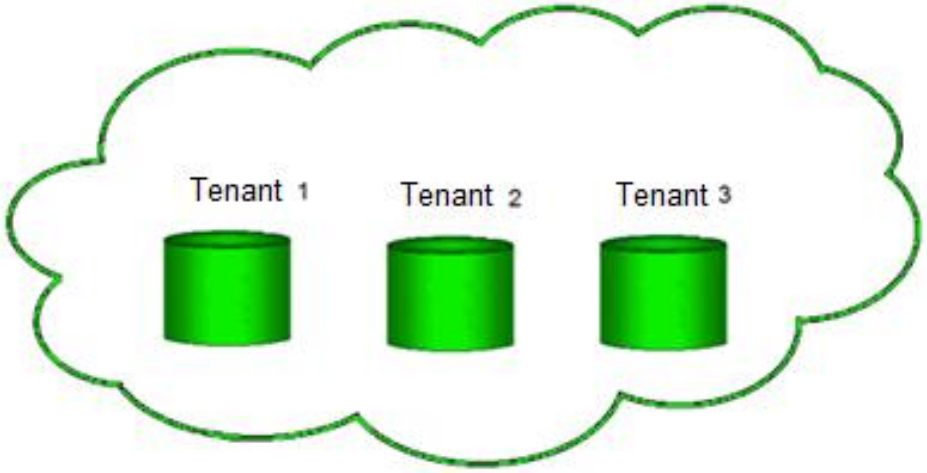


Fig. 1.Single tenant, separate data bases

2.MultiTenant with identical schemas - this approach provides scalability but limits configuration options for each individual tenant, forcing them to use only the already defined schemas.

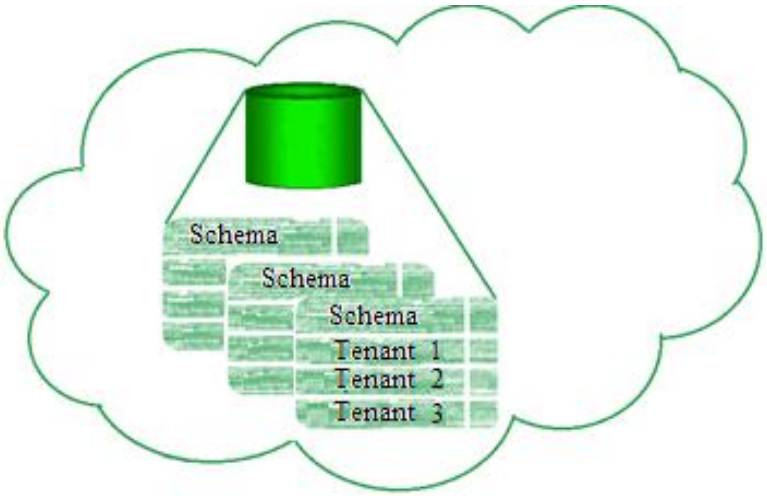


Fig.2 MultiTenant with shared data base, identical schemas

3. MultiTenant utilizing different tenant schemas - this approach provides a wide range of options for configuring custom schemas but limits the supplier's ability to maintain a basic program, thus it is forced to provide a number of specific applications within the core code.

Fig.3 MultiTenant with shared database, different schemas

Bulgarian companies that provide Software as a Service often use this option with an addition thereto which facilitates their work and somewhat mitigates the negative aspects of the approach. Companies typically maintain a basic system where they introduce all corrections and errors encountered during testing and operation of all systems. For each new tenant is made a copy of this basic system which is further on adjusted, the schema is complemented and pre-programmed if necessary - as required by the given object.

Multitenancy in the three cloud deployment models

Multitenancy occurs specifically in all three cloud deployment models:

Public cloud. In this model the delivery of services is performed off-site by third-party service provider which shares resources in multitenant operating environment. Physical infrastructure is owned and managed by the service provider.

Private cloud. In this model the service is provided by an organization or its designated service provider as the services are available in single-tenant mode. Physical infrastructure can be owned and managed by the organization or by given service provider.

Hybrid cloud. In this model the service is composed of two or more clouds (private or public) that remain unique units but are bound together by standardized technologies that allow data and applications transferability.

First we will consider the case defined as private cloud (Fig. 3). In this case the organization that owns the cloud may have different units or different sub-units within a single unit using the cloud. Each of them can have different levels of security and access to various services. The cloud comprises of different virtual machines (VM), which are managed by managers of virtual machine (MVM).

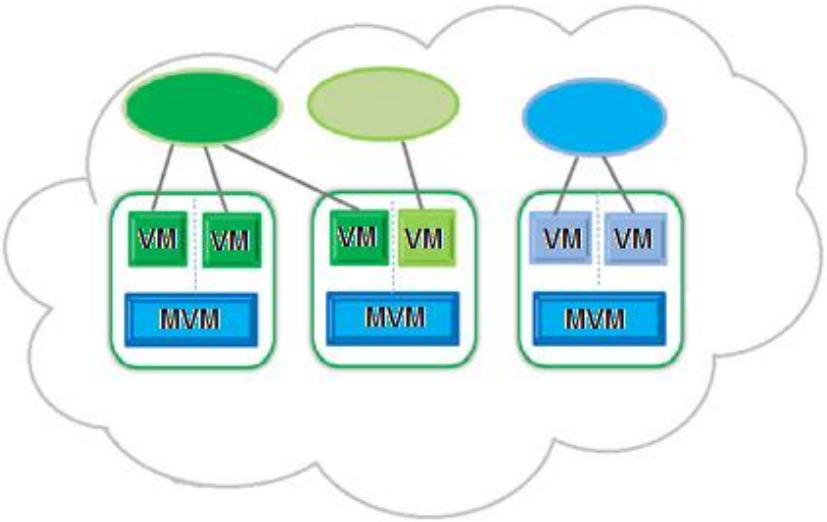


Fig.4. MultiTenant - private cloud

In the public cloud different organizations and separate tenants can use the same services provided by the supplier sharing common virtual space. It provides high level of accessibility, operational efficiency, economies of scale and cost savings.

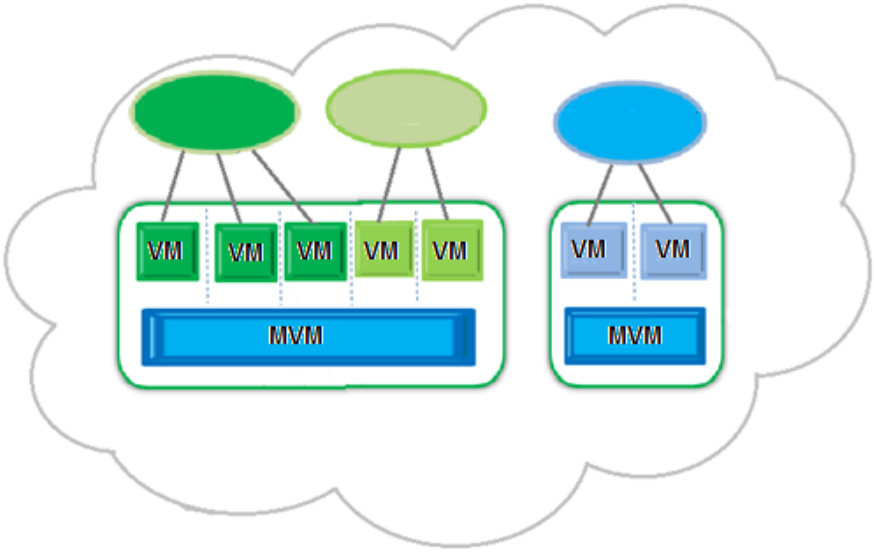


Fig. 5. MultiTenant - public cloud

What is typical for the Bulgarian companies? In our country seems to dominate Private cloud. Organizations that hire Software as a Service install it on their own cloud and use it in single-tenant mode.

Summary

The report treats only a fraction of the applications of Multitenancy in cloud computing. Multitenancy demonstrates itself in different ways depending on the model of service provision and of the deployment model of cloud computing.

Currently in Bulgaria cloud computing seems to be used mostly in advertising and education. Companies use clouds as a "show room" - to demonstrate the software and platforms they offer. Tenants are offered a free of charge alternative in mode MultiTenant to exploit the capabilities of the software products that developers offer. After the expiry of a fixed period of time they will be able to purchase the real version of the product, if the user wishes the product may not be provided as a cloud computing and as a traditional software.

Literature:

1. Велев Д., Кисимов В., Методи за проектиране на бизнес информационни системи на базата на Web 2.0 и Cloud Computing, Научни трудове на УНСС, том 2, 2009, (209–246), ISSN: 0861-9344.

2. Велев Д., Приложни аспекти на облачните изчисления в бизнеса, Изд. „Авангард Прима“, София, 2012, 137 стр., ISBN: 978-619-160-051-9.

3. Майкъл Кьоглер, За ползите от облачните услуги, iNEWS http://inews.bg/%D0%97%D0%B0-%D0%BF%D0%BE%D0%BB%D0%B7%D0%B8%D1%82%D0%B5-%D0%BE%D1%82-%D0%BE%D0%B1%D0%BB%D0%B0%D1%87%D0%BD%D0%B8%D1%82%D0%B5-%D1%83%D1%81%D0%BB%D1%83%D0%B3%D0%B8_l.a_i.112880.html

4. Пазарът на облачни технологии еволюира от капацитет към потенциал. http://computerworld.bg/cloud/1/

Cloud Business Intelligence Services for Small Business

Anna Yordanova

Abstract. The economic crisis is requiring fast development and increasing company's efficiency. More than ever the quality and speed of the business decisions has to be improved. The cloud business intelligence services can provide the tools which will bring the competitive advantage to the business. Cloud is a computing model providing web-based software, middleware and computing resources on demand. Cloud computing can allow the users to access the latest software and infrastructure offerings to foster business innovation. Cloud computing helps businesses transform their operations and technology by establishing a flexible, adaptable IT environment to quickly meet changing requirements. The need to accomplish more with less makes the cloud's on-demand nature perfect IT resource for organizations that need to operate within a tight budget but also need to expand their capacity as their business grows. This paper will discuss the topic of Cloud Business Intelligence service model in more detail, emphasizing on the services for small business. The presented model is putting stress on the following aspects of the Cloud BI for small business: cost effective, easy to be used, well performing, visual and secure on-demand BI solution. An overall comparison of the existing cloud BI solutions will be presented.

Keywords: Cloud Computing, Business Intelligence, Services Oriented Architecture, Software as a service, Small Business.

1 Introduction

It has become apparent in recent years, that small and medium-sized businesses have not been able to afford the high cost and complexity of Business Intelligence, while larger business entities became stuck with clunky, brittle and unwieldy BI platforms and architectures that never delivered their promised ROI (Return On Investment) [16]. Therefore, a scalable, cost effective and easy entry solution is needed. Cloud is a computing model providing web-based software, middleware and computing resources on demand. Cloud computing can also go beyond cost savings by allowing the users to access the latest software and infrastructure offerings to foster business innovation.[1] Cloud computing helps businesses of all sizes transform their operations and technology by establishing a flexible, adaptable IT environment to quickly meet changing requirements. The need to accomplish more with less is at the financial core of every medium sized business, and the cloud's on-demand nature makes it the perfect IT resource for organizations that need to operate within a tight budget but also need to expand their capacity as their business grows [2]. The cloud offers access to data, applications and other resources without the need for program installation. This equals major convenience when doing work on a portable device like a laptop, tablet PC or smart phone. The devices are not only free from the clutter of numerous installs, but the company's IT team isn't bogged down with installing, reinstalling, and troubleshooting numerous devices for each employee [15]. On-demand Business Intelligence means on-demand licensing; this result in tremendous cost savings for companies that have

for years squandered money on software and licenses that filled a niche role or were not as heavily used as anticipated [16].

2 MOTIVATION

Cloud Business Intelligence is gaining ground, according to a recent survey of BI Leadership Forum, a global network of BI Directors and other BI professionals (www.bileadersship.com). More than one-third of organizations are currently using the Cloud for some part of their BI program, according to the survey (See figure 1) [20].

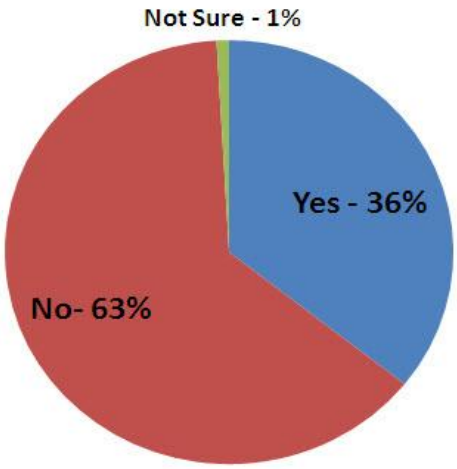


Fig. 1. Are you using the Cloud for any part of your BI program? Source: BI Leadership Forum, June, 2011. Based on 112 responses. www.bileadership.com

The organizations that have embraced the Cloud point to "speed of deployment" (30%) and "reduced maintenance" (30%), followed by "flexibility" (19%) and "cost" (11%) (See figure 2) [20].

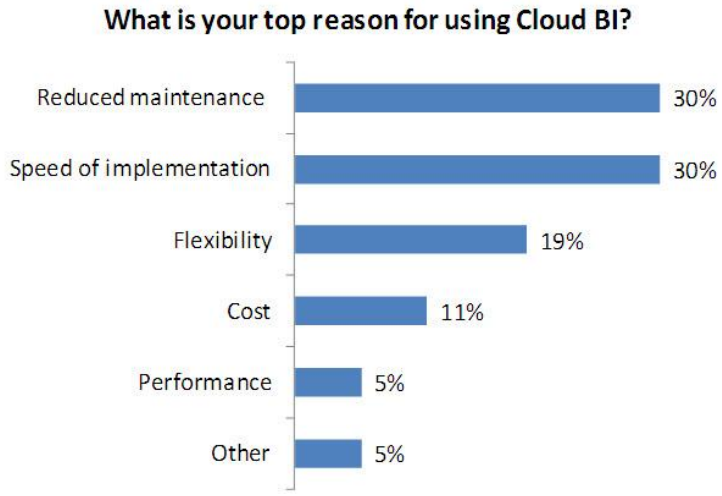


Fig. 2. Motivating Factors

Almost two-thirds of the Cloud BI users (65%) said they plan to increase their usage of Cloud BI in the next 12 months. Only 3% said they would decrease usage while another 16% will keep their implementation the same and 16% weren't sure. (See figure 3.)

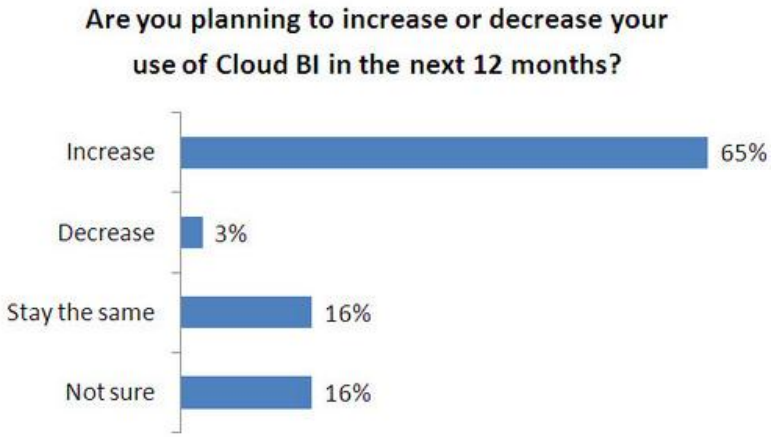


Fig. 3. Future Usage

Among respondents who are not using Cloud BI, 16% said they plan to implement Cloud BI in the next 12 months and 32% were not sure. However, it may take a five to 10 years for Cloud BI to reach the tipping point where it becomes a mainstream component of every BI program. Given Cloud BI's benefits, this trajectory is inevitable [20]. According to Wayne Eckerson "Small Companies Lead the Way". A closer look at the data confirms what many pundits have said about the target market for Cloud-BI software: that it's currently ideal for small companies with few IT resources, limited capital to spend on servers and

software, and minimal to no BI expertise. Almost half of small companies under $100M in annual revenues (46%) use Cloud BI in some shape or form. In contrast, large companies with over $1B in annual revenues are almost less than half as likely to adopt the cloud (29%), while medium-sized companies with between $100M and $1B in annual revenues lag further behind with less than one-fifth using BI in the Cloud (18%). (See figure 4.)

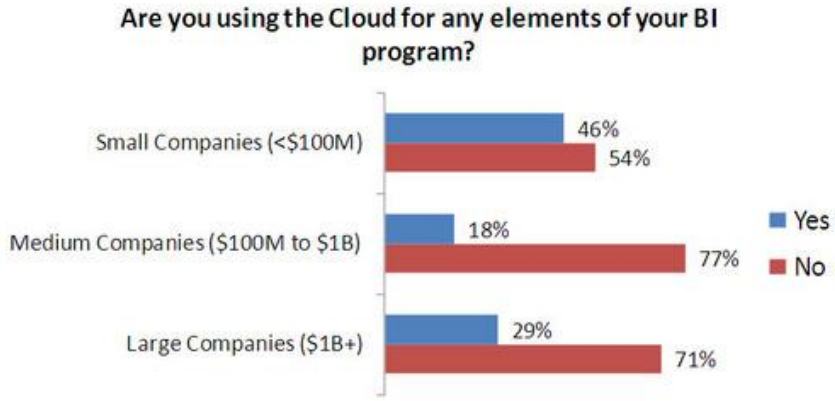


Fig. 4. Cloud BI Deployment by Company Size

For small businesses without legacy BI applications, Cloud BI services are a godsend. The economics and convenience are compelling. Instead of passing around spreadsheets, small companies can implement a Cloud BI service to standardize reports and dashboards and make them available to all employees anywhere via a Web browser. Wayne Deer, vice president of operations, at Gazelle, an electronics recycler, which uses GoodData's Cloud BI service says: "What's refreshing for me is that I can go in at any time of day and [run a] report on any metric in our organization, such as item received delivered, inspected at the category, personnel, or employee level and track it by any time period,"

3 CHARACTERISTICS OF THE POSSIBLE SOLUTION

It has become apparent in recent years that to accomplish more with fewer resources is an issue that many small businesses feel strongly about. Because of this reason a research will be made. The goal of this research is to find out whether a solution with the outlined below hypothetic characteristics already exists.

Hypothetic Characteristics of Cloud Business Intelligence Software as a Service.

- cost effective (< $100 per month)
- easy to be used (intuitive navigation)
- well performing (returning a result for less than 30 sec –assumption: a network speed is not a bottleneck)
- visual and secure on-demand BI solution
- easy upload of data (upload of xls files, connection to data sources, social media, other)

- Size of the business < 15 employees

The solution can be beneficial for the users by providing:

- IT services – maintenance and support of the application
- Business know-how - predefined reports and help with tips for the analysis

Potential target group of the proposed solution can be small and medium business.

4 PROBLEM STATEMENT

Identify whether a SaaS Business Intelligence solutions for small business are already build. If such solutions exist identify whether a solution with the above characteristics exist.

5 APPROACH

Before starting the market research two terms has to be clarified - "Cloud computing" and "Small Business". Definition of Cloud computing – to access data and applications without on-site infrastructure (in the cloud).

- data processing,
- storage and backup,
- maintenance,
- administration and troubleshooting

are all taken care of by the service provider.

Definition of Small Business:

- "Small" in terms of government support and tax policy varies by country, by industry and ranging
- <15 employees under the Australian Fair Work Act 2009,
- < 50 employees according to the definition used by the European Union, and
- < 500 employees in U.S.

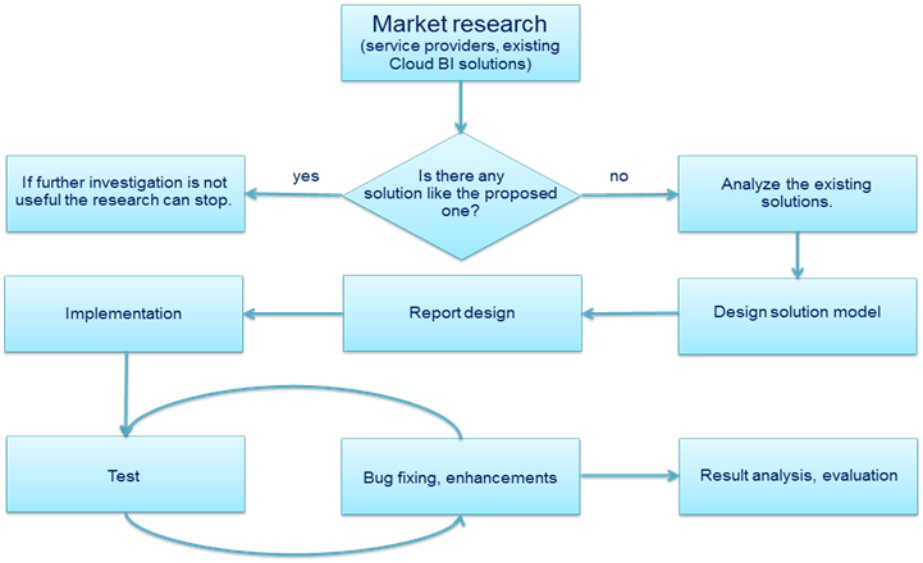
In this research "small business" is considered business with less than 15 employees

5.1 RESEARCH, SYSTEM DESIGN AND TECHNOLOGY

The focus of the paper is to present current result of the conducted market research but the "market research" itself has to show whether further research and development of such application (like the one described above with hypothetic characteristics) is making sense.

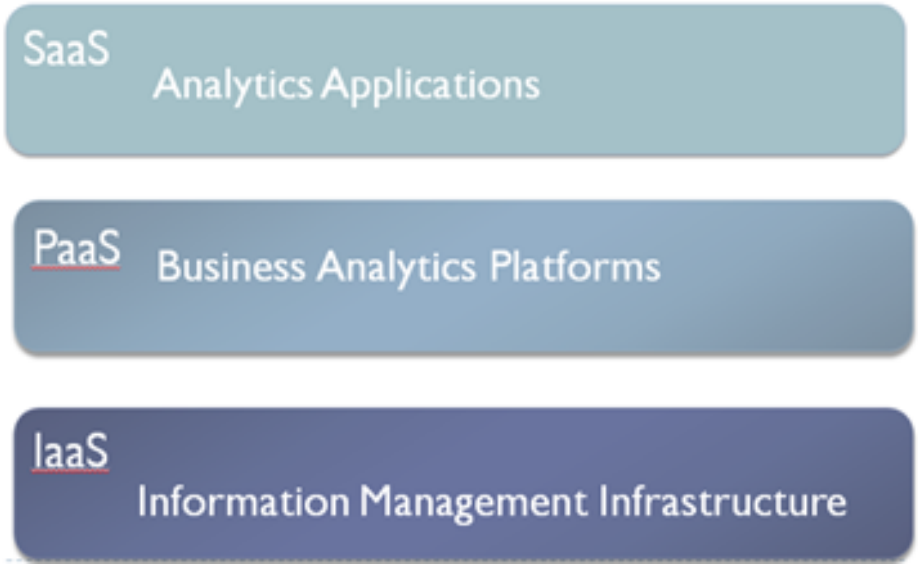
If the result is showing that there are already many similar solutions then future research focus might be changed. In case there are no similar solutions then a development work might start. It is very important to be analyzed what is currently offered by the service providers. When the market is analyzed the "design solution model" can be clarified. The "report design" phase can be quite challenging because the design itself is a key success factor and has to consider many aspect in relation with the fact – variety of companies has to

be able to use the solution. A good report design can bring higher business value to the companies. After completing the implementation further testing and bug fixing will be needed. When the solution is at a mature stage the development can stop. The last step will be to analyze the result and evaluate whether the created solution is useful for small companies. Depend on the results of the evaluation a second phase of development can start.



5.2 MARKET RESEARCH

According to John Hagerty from Garther group the following Mapping of Business Analytics to Cloud Services is given:



On the market there are many service providers which offer Infrastructure as a Service, Platform as a Service and Software as a Service. On the figure below are given some of the service providers which offer different cloud solutions. The companies are spited by layers but it is important to be highlighted that they are expanding their business and service packages really fast and soon they will probably offer solutions in the 3 different layers.

2012 Best Cloud Computing Services Comparisons and Review

Fig. 5. Cloud service providers

According to [26] while there is certainly some overlap between each of these services, cloud services are unique because they allow you to view, edit and share files saved in the cloud. Even though no two cloud services are identical, each of the reviewed services provides the same fundamental features and functionality.

Features

The most important factor to consider in a cloud service is features, including the type of content you can store. The best cloud computing services are those that allow you to upload and save any type of file you would save on your local hard drive, from word documents to music files and everything in between. Any cloud service you consider should also allow you to view, edit and share your content regardless of what computer or device you are using. Other features to look for include automatic syncing of your files across all your devices, and password-protected sharing and file encryption to safeguard your content.

Mobile Access

Arguably one of the biggest selling points of cloud computing services is their wide-ranging access. Whether you're on your work computer at the office or at home on your iPad, cloud services allow you to access your content anywhere, anytime and on any one of your devices. Look for services that offer the greatest range of mobile access, including apps for popular smartphones and the ability to log into your account from any mobile browser.

Ease of Use

Considering how often you'll likely be accessing your content in the cloud, it's important to select a cloud service that is intuitive and straightforward. The service's interface and tools should be easy to navigate and convenient to use.

Help & Support

Getting help when you need it is crucial when using any type of technology, including cloud services. Available support options should include technical assistance via telephone, email and live chat. The service should also provide a knowledgebase and user forums as resources.

2012 Best Cloud Computing Services Comparisons and Review

Rank	#1	#2	#3	#4	#5	#6
10-9 Excellent 8-6 Good 5-4 Average 3-2 Poor 1-0 Bad	iCloud	Egnyte HybridCloud	Google Apps	OpenDrive	Dropbox	Amazon Cloud Drive
Ratings	**9.70**	**9.53**	**9.38**	**9.23**	**8.90**	**7.98**
Overall Rating Features Mobile Access Ease of Use Help & Support	10 9 8 7 6 5 4 3 2 1					
Price Comparison						
Lowest Monthly Cost	$1.66	$24.99	$5	$5	$9.99	$1.66
Storage Space	15GB	150GB	1GB, 25GB for email	100GB	50GB	20GB
Content						
Email	✓	✓	✓			
Contacts	✓	✓	✓			
Calendar	✓	✓	✓			
Documents	✓	✓	✓	✓	✓	✓
Spreadsheets	✓	✓	✓	✓	✓	✓
Music	✓	✓	✓	✓	✓	✓
Photos	✓	✓	✓	✓	✓	✓
Videos	✓	✓	✓	✓	✓	✓
Slideshows	✓	✓	✓		✓	✓
Features						
View Files	✓	✓	✓	✓	✓	✓
Edit Files	✓	✓	✓	✓	✓	
Automatic Sync	✓	✓		✓	✓	
Password Protection	✓	✓	✓	✓	✓	✓
File Encryption	✓	✓	✓	✓	✓	✓
Media Streaming	✓			✓		✓
Share Files		✓	✓	✓	✓	
Desktop App/Access		✓		✓	✓	
Mobile Access						
iPhone	✓	✓	✓	✓	✓	
iPad	✓	✓	✓	✓	✓	
Android		✓	✓		✓	✓
BlackBerry			✓		✓	
Windows Mobile			✓			
Mobile Browser		✓	✓	✓	✓	✓
Help & Support						
Phone Support	✓	✓	✓			✓
Email Support	✓	✓	✓	✓		✓
User Forums	✓	✓	✓	✓	✓	✓
Knowledgebase	✓	✓	✓	✓	✓	✓
Online Ticket Form	✓	✓		✓	✓	
Live Chat			✓			
Supported Configurations						
Windows 7	✓	✓	✓	✓	✓	✓
Windows Vista	✓	✓	✓	✓	✓	✓
Windows XP	✓	✓	✓	✓	✓	✓
Mac OS X	✓	✓	✓	✓	✓	✓
Linux		✓			✓	

Fig. 6. 2013 Best Cloud Computing Services Comparisons and Review [26]

5.3 CURRENT OBSERVATIONS

- A Saas Business Intelligence solution for small business with the above parameters is not created yet according to the current research analysis state.
- A solution as Google Analytics is closer to the analyzed solution but the given analisys is about web analytics and not business analysis.

- SAP BI OnDemand is easy to be used, but the cost is not yet clear. Might be suitable technology for the application development.

6 RESEARCH QUESTIONS AND FUTURE STEPS

Some of the arising questions:

- Which are the right tools for the small business?
- What analyses are with high priority for the small business?
- What price the small business can afford?
- What is a small business? (Different definitions in the different countries)
- What cost the small business afford?
- How the small business can handle the data preparation?
- How easy can be the data upload?
- How detailed categorization of the business can be used?

7 CONCLUSIONS

The use of Cloud Business Intelligence Software as a Service is just starting to develop and expand. SaaS Business Intelligence for small business can be compared with e-commerce in terms of trust and security. Before few years nobody believed that e-commerce will become so widly used and nowadays is just part from our lives. Even thought many questions arrise and further analysis of the existing solutions and "small business" requirements is needed before startig a design phase the author believe that Cloud Business Intelligence Software as a Service has the potential to growth.

References

[1] IBM cloud computing:Rethink IT. Reinvent business, http://www.ibm.com/cloudcomputing/us/en/
[2] Transform your business with IBM cloud capabilities, http://www.ibm.com/midmarket/us/en/cloudcomputing.html
[3] The Cloud and SOA - Creating an Architecture for Today and for the Future, Campgemini and HP (2008), http://www.hp.com/hpinfo/analystrelations/wp_cloudcomputing_soa_capgemini_hp.pdf
[4] Manan, G., Cloud ready business intelligence with Oracle Business Intelligence 11g, An Oracle White Paper (2010)
[5] Berkowitz J. (2009), Cloud Computing (Part 1) : Advantages, Types and Challenges, http://crmweblog.crmmastery.com/2009/11/cloud-computing-part1-advantages-types-and-challenges/
[6] Anchan P. (2012), 3 Major Types of cloud computing services, http://www.cloudcomputingpath.com/3-major-types-of-cloud-computing-services/
[7] Dave Harfield (2012), The cloud computing book
[8] Working smarter with a Smart SOA foundation (2010), ftp://public.dhe.ibm.com/software/solutions/pdfs/7946_Working_smarter_Brochure.pdf
[9] Dimitrov V. (2009), Service Oriented Architecture

[10] Butner K. (2012), Combating risk with predictive intelligence, http://public.dhe.ibm.com/common/ssi/ecm/en/gbe03481usen/GBE03481USEN.PDF
[11] Mircea M. (2008), Combining business intelligence with cloud computing to delivery agility in actual economy, http://pits-bi-real-time.googlecode.com/svn/trunk/Tesis/Lecturas/Real-time/COMBINING%20BUSINESS%20INTELLIGENCE%20WITH%20CLOUD%20COMPUTING%20TO%20DELIVERY%20AGILITY%20IN%20ACTUAL.0ECONOMY.pdf
[12] Institute for Business Value IBM, http://www-935.ibm.com/services/us/gbs/thoughtleadership/ (http://public.dhe.ibm.com/common/ssi/ecm/en/gbe03482usen/GBE03482USEN.PDF)
[13] Kobielus J. (2009), Appliance Power: Crunching Data Warehousing Workloads Faster and Cheaper Than Ever, (http://www.cio.com/documents/whitepapers/ibm_swg_appliancepower.pdf)
[14] Top 10 Cloud computing companies in the world (2012), http://www.cloudcomputingpath.com/top-10-cloud-computing-companies-in-the-world/
[15] Dieterle M.(2011), BI in the Cloud – Why Make the Move?, http://biblog.arcplan.com/2011/07/bi-in-the-cloud-%e2%80%93-why-make-the-move/
[16] Laurent W. (2009), Business Intelligence In The Cloud - BI And Software As A Service
http://www.dashboardinsight.com/articles/new-concepts-in-business-intelligence/business-intelligence-in-the-cloud.aspx
[17] McKendrick J., Fastest-Growing Category of Cloud Computing: Business Intelligence and Analytics, http://www.forbes.com/sites/joemckendrick/2012/07/19/fastest-growing-category-of-cloud-computing-business-intelligence-and-analytics/
[18] Trajman O., Business Intelligence in the Clouds (2009) , http://www.information-management.com/infodirect/2009_111/10015046-1.html
[19] Burbach S., Birst Ranks First Among Cloud BI Vendors in Cloud Business Intelligence Market Study, http://www.birst.com/company/press/birst-ranks-first-among-cloud-bi-vendors-cloud-business-intelligence-market-study
[20] Eckerson W., Cloud BI Adoption: Gauging Market Demand (2011), http://www.b-eye-network.com/blogs/eckerson/archives/2011/09/cloud_bi_adopti.php
[21] Eckerson W., What is Cloud Computing? (2011), http://www.b-eye-network.com/blogs/eckerson/archives/2011/07/what_is_cloud_c.php#
[22] Eckerson W., Expectations Vs Reality in the Cloud: Understanding the Dynamics of the SaaS BI Market, http://www.b-eye-network.com/blogs/eckerson/archives/2011/07/expectations_vs.php
[23] Eckerson W., Jockeying for Position in the Analytic Platform Market, http://www.b-eye-network.com/blogs/eckerson/archives/2011/03/jockeying_for_p.php
[24] Hagans Ch., Business Intelligence in the Cloud (2011), http://cloudcomputingtopics.com/2011/09/business-intelligence-in-the-cloud/
[25] Frank G., BI & DR In The Cloud: A Smart Alternative For Your Business?, http://www.webroot.com/En_US/business/articles/bi-dr-in-the-cloud-a-smart-alternative-for-your-business
[26] 2012 Best Cloud Computing Services Comparisons and Review, http://cloud-services-review.toptenreviews.com/

Business Process Management and Web 2.0

Violeta Kraeva, Petja Emilova

D. A. Tsenov Academy of Economics – Svishtov, Bulgaria
e-mail: kraeva@uni-svishtov.bg; petiaem@uni-svishtov.bg

Abstract. New social network capabilities will determine the next generation of applications for Business Process Management (BPM). Experts at Forrester announce about the emergence of Social BPM technologies, such as process mashups, process wikis, and BPM-as-a-Service. Some communities are formed around the business processes that extend to different functional units, intersecting across departmental value chains, partners, suppliers and customers – a whole BPM society that can now benefit from solutions built on the basis of social networks. BPM provides the context of cooperation and interaction of network partners and social networks support activities that make up business processes and improve the life cycle of an individual process. Several categories of BPM societies are formed: Society of BPM project within the enterprise; society of partners in the supply chain; research community to discuss BPM standards, best practices, methodologies and models; societies for discussion, bloggers known as BPM bloggers; communities focused on specific domain - telecommunications, financial services, manufacturing, etc. The purpose of this report is to investigate characteristics and trends in use of social networks for the realization of cooperation underlying the BPM.

Key words: BPM, Web 2.0, social networks, collaboration.

1. Introduction

The social media are rapidly making their way through and acquire a greater significance in the way the business is done. The latest researches show, that a huge part of the Bulgarian companies account for the increasing significance of the social networks for the improvement of their competitiveness and stimulation of innovation. The number of registered users in one of the biggest business social networks LinkedIn is constantly increasing, which is an expressive evidence for the huge interest and the capabilities for the business. During the first quarter of 2012, the most popular social network Facebook reached 835 million users, which makes 12.1% from the world population[8]. This led to increased attention and focused effort for the transformation of this virtual social environment into an effective business environment. Currently are developed more closed social networks, narrowed by professional interests.

The purpose of this report is to investigate the special features and trends in the use of the social networks for the realization of business partnership, based on the BPM.

[8] Internet World Status, Usage and Population Statistics. http://www.internetworldstats.com/stats4.htm

2. Business processes and social media

2.1. Business processes

It has become a tradition, the technologies for the management of the business processes (BPM) to be used by the businesses for the improvement of efficiency, reducing the production cycles, cutting the costs and increasing the quality of the goods and services. BPM presumes the development of strategic approach, pointed toward the congruence of all processes in the boundaries of one organization in accordance with the goal to deliver value to the customer. It encourages the improvement of the business results, along with the aspiration for innovation, flexibility and integration of the technologies.

Structured and nonstructured BP. The business processes could be categorized as structured and nonstructured processes[9]. The structured processes are repetitive and based on a clearly defined work flow. In this case the process is routine, the flow is clearly determined, and even if in the flow of the process, the information is changed, the defined flow of activities is not changed. The nonstructured processes are more complicated, partially defined, and for the management of particular work flow must be taken in consideration a lot of factors, such as: the possibility of the data to vary; the probability the number of participants to vary; to what extend is known the number of the steps forming the process (apart from in the start and ending points). The work flow, based on these factors is very dynamic. As the nonstructured processes are flexible and based on decisions, which have to be taken (in the most common case decisions, taken in cooperation), often *in them take part the people and form significant influence* over the way the process is conducted.

2.2. Web 2.0

There are various interpretations for the name Web 2.0. According to one of them "2.0" means the second generation of World Wide Web. The first generation of Web, was focused over relatively static Web sites and Web presence. The new generation of Web secures a much richer experience and more important – focuses over the *society*. Wikis and blogs are probably the most popular social network applications. The blogs allow the author the share ideas or comments on certain matters. Wikis give the mechanism for cooperation in the process of authorship. There are other applications, namely – discussion themes, inquiries, instant messages (IM), shared white boards, common bulletins and editing in cooperation. *The purpose of Web 2.0 is to connect people, form societies and encourage the common development in Web.* As a result of that, one of the biggest profits of Web 2.0 is expected to be realized through BPM Suite, namely – partnership across the business organization and between trade partners in Internet.

Web 2.0 technologies develop BPM **beyond the fortification wall**, using technologies as cloud computing, BPM-as-a-Service and software, offering social functionality for the support of the business processes.

BPM-as-a-Service allows the business users to model, develop and realize the management of the process in the cloud. The organizations receive the opportunity for the control over the processes, cutting significantly the costs and IT requirements towards the infrastructure. The different suggestions, which recently showed up in the **cloud services,**

[9] Jenkins, T. Managing Content in the Cloud. Second Printing, October 2011, p. 186.

vary in wide boundaries – from hosting of modeling of the processes in the cloud to vertical or horizontal decisions for the work flow. The organizations are using them by the principle – "we pay only for that, what is used".

Taking part in complex business processes the people make variety of interactions with other individuals, which turn BPM into a social application. As BPM includes the cloud, the capabilities of the social networks find place in many of the BPM applications.

As the short period of skepticism and mistrust elapsed, during which the use of social networks and other Web 2.0 technologies in the enterprise was perceived as distracting the attention of the personnel from the production process and job responsibilities, it came time for the real appraisal of their potential for the business and they are already turning into major business instruments for work in net, communications and marketing. The interaction between structured processes, such as workflow and nonstructured processes, such as interaction in the social networks are emerging quickly, integrating instant messages, wikis, discussion forums and cooperation with BPM.

As with every beginning the threats and abilities for abuse of Web 2.0 instruments are huge (created blogs and wikis with doubtful quality for example). Nevertheless the social networks and other Web 2.0 instruments are turning into a major factor in the enterprise. These instruments are gradually imposing themselves as major components in the portals of the enterprise. The corporate portals are evolving to an environment, where the employees could communicate and cooperate through discussion themes, to express opinions on different matters in the blogs and to cooperate in the creation of wikis on professional matters. In this way more and more organizations encourage and stimulate the internal communication **(beyond the defense wall)**, developing on the basis of social networks.

2.3. Cooperation

The effective interface of the BPM applications with applications such as ERP, SCM and CRM allows the organizations to speed up the workflows and to carry out transactions in the boundaries of several applications. BPM however, assumes *the management of the interaction between the people* along the workflows.

The cooperation in Groupware and Work Group computing has two dimensions: **time** and **place**[10]. In regard with time we distinguish *synchronous networking and collaboration*, which is realized in one and the same time and a*synchronous networking and collaboration*, in which the parties take part in the process of cooperation in different time. In regard with the time, we have cooperation *in one and the same place* or *in different places*. Thus could be outlined four groups of instruments for cooperation, as shown in table 1.

[10] Khoshafian, Setrag. MyBPM: Social Networking for Business Process Management. Pegasystems Inc., http://www.mitcio.com/files/sponsorwp/PegaSystems-MyBPM_Social_Networking_for_BPM-2010.pdf, p.4.

Table 1. Dimensions of cooperation

	Same Place	***Different Places***
Synchronous Networking and Collaboration (same time)	• Computer-enabled Meeting Rooms • Virtual Meeting Rooms (e.g. Second Life)	• Chat and/or Instant Messaging • Electronic Meetings: Videoconferencing, • Web Meetings • Shared White boards • Shared Applications
Asynchronous Networking and Collaboration (different time)	• Walls (e.g. on Facebooks) • Discussion Threads • Virtual Rooms • Kiosks • Electronic Bulletin Boards • Blogs and Wikis	• E-Mail • Workflow • Task Lists • Collaborative Document Authoring • Shared Calendaring • Surveys and Voting

Source: Dr. Setrag Khoshafian. MyBPM: Social Networking for Business Process Management. Pegasystems Inc., http://www.mitcio.com/files/sponsorwp/PegaSystems-MyBPM_Social_Networking_for_BPM-2010.pdf, p.4.

The specialists scrutinize the collaboration, relating it to the stages of the life cycle of BPM, which are: - modeling of the process applications, deploying the application, the execution of the process, as well as analysing process performance. For the consecutive and constant improvement of the execution of the processes a monitoring is undertaken, analysis and optimisation of the processes. These are activities, the effectiveness and efficiency of which directly depends, from the communication and collaboration of the participants in the business process.

2.4 Social BPM

In the process of collaboration, taking place in the social media is forming several categories of BPM societies[11]:

- BPM Projects Society Within The Enterprise. During the construction, deployment and analysis of the implementation of the BPM applications are developed social relationships among the different participants in each of the phases of the life cycle. For middle and upper tier organizations, the processes and respectively the relationships, related with the BPM project could include or cross the internal value chains and business functions. The sustainment of collaboration via social networks among the functional units of the organizations is a bigger challenge than the collaboration in the boundaries in a separate structural unit. This is collaboration in the boundaries of an enterprise.
- BPM Society Across Trading-Partner Value Chain. The enterprises are interacting and exchanging experience with suppliers and clients. The supply chain includes

[11] Khoshafian, Setrag. MyBPM: Social Networking for Business Process Management. Pegasystems Inc., http://www.mitcio.com/files/sponsorwp/PegaSystems-MyBPM_Social_Networking_for_BPM-2010.pdf, p. 6.

several organizations. This represents the next level of BPM collaboration and includes the participants from the enterprise along the value chain. Here in the social networks are forming and sustaining the B2B societies and the extended business process of the organization includes also the partner system.

- BPM Communities. The expectations are these communities to impose as networks of BPM standards, best practices, methodologies and templates. There do already exist several temporary discussion forums and research BPM societies, as well as BPM bloggers. Despite that, the specialists consider, that the greater value will be achieved, when the community is focused over given vertical domain (for example production, telecommunication, financial services, insurance, healthcare, transport, education, government and others.)

Everything in a given enterprise is done in the context of a given process. The process includes policies and procedures, as most of them are not modeled and automated yet. The purpose of every collaboration, regardless of whether it comes for innovation, development of products and services, marketing, sales, production analysis, contracting, supply, building of partnerships or processing of the clients requests, is to support the business policy or procedures, i.e. the business processes. In this regard is emerging a powerful collaboration between BPM packages and social networks. According to Khoshafian[12], ***"Business processes provide the context of collaboration, and social networking supports and augments the various phases of the BPM continuous improvement lifecycle."***

The author systemizes the instruments of the social networks, which are used together with the processes in the following way:

- *Use of* Blogs and Wikis for Complex Compliance Applications. The instruments for collaboration in the social networks could be used, for the clarification and discussion of the best practices for the attainment of compliance with legal requirements and control over their application. Discussion forums could be made, blogs and wikis for clarification and discussion of the specific for the business processes flows, rules, methodologies, possible optimizations and practices, which automate the processes and the compliance rules.
- Synchronous Collaboration within Process Execution), which if in the form of meetings in the process of negotiation before the contract conclusion, would suggest:
  - Scheduling Collaboration Session.
  - Synchronous Collaboration Session.
  - The Session.
- *Discussion threads, surveys, and blogs when building a BPM application.*

3. Conclusion

We are witnessing a process of powerful synergy between BPM and social media. BPM society of users includes different functional units, crosses departmental value chains, trade partners and common BPM societies. The collaboration in the social networks could be organized in compliance with the dimension - time (synchronous vs. asynchronous) and dimension – place (one and the same and different places).

[12] Пак там.

Every aspect of the BPM collaboration could use synchronous or asynchronous the social networks. So, as long as the business processes provide the context of collaboration, the social networks maintain and ameliorate the different activities for continuous improvement of the life cycle of the BPM applications.

References

1. Jenkins, T. Managing Content in the Cloud. Second Printing, October 2011

2. Khoshafian, Setrag. MyBPM: Social Networking for Business Process Management. Pegasystems Inc., http://www.mitcio.com/files/sponsorwp/PegaSystems-MyBPM_Social_Networking_for_BPM-2010.pdf

3. Internet Word Status, Usage and Population Statistics. http://www.internetworldstats.com/stats4.htm

TRENDS IN THE USE OF WEB SERVER SOFTWARE IN BULGARIAN BANKS

Pavel Petrov

Faculty of Informatics
University of Economics - Varna
9002 Varna, 77 Kniaz Boris I, Blvd.
Phone: ++359 52 660412
E-mail: petrov@ue-varna.bg

Abstract. The publication presents the results of empirical research on web server software used by Bulgarian banks in 2012 and compared to similar survey conducted in 2009. Studied were 27 banks. Survey reveals some interesting details about the widespread use of free software from Bulgarian banks - which is in generally conservative institutions that have significant financial resources and for which safety and security are among the most important priorities.

Keywords: web server, banks, software.

Introduction

The purpose of this publication is to summarize the information gathered in the course of research about web server software used by Bulgarian banks in 2009 and 2012. The scope of the study is limited to the web server serving the main site of one specific bank and was done in the period May-August 2009 and in the month of July 2012. Some bank's branches have not local website in Bulgarian language and were not included in the list.

In most cases, the responsible for planning, managing and configuration of computer network is system administrator. Daily he monitors systems to ensure proper operation and to prevent unauthorized access or attacks. In this sense, the configuration of what information will provide the web server is his responsibility.

Bulgarian banks websites

The list of banks authorized to operate in Bulgaria and branches of foreign banks operated in Bulgaria is taken from the Bulgarian National Bank (The Central Bank of the Republic of Bulgaria) website [1].

Table 6. Banks and branches of foreign banks in Bulgaria and their addresses surveyed in 2009 and 2012[13]

No	Bank/Branch	Address
1	Allianz Bank Bulgaria AD	bank.allianz.bg
2	ALPHA Bank S.A. – Bulgaria Branch	www.alphabank.bg
3	BNP Paribas S.A. - Sofia Branch	www.bnpparibas.bg
4	Bulgarian Development Bank AD	www.nasbank.bg
5	Bulgarian-American Credit Bank AD	www.bacb.bg
6	Central Cooperative Bank AD	www.ccbank.bg
7	CIBANK JSC	www.cibank.bg[14]
8	Corporate Commercial Bank AD	www.corpbank.bg
9	D Commerce Bank AD	www.dbank.bg
10	DSK Bank EAD	www.dskbank.bg
11	Emporiki Bank - Bulgaria EAD	www.emporiki.bg
12	Eurobank EFG Bulgaria AD	www.postbank.bg
13	First Investment Bank AD	www.fibank.bg
14	ING Bank N.V. – Sofia Branch	www.ing.bg
15	International Asset Bank AD	www.iabank.bg
16	Investbank AD	www.ibank.bg
17	MKB Unionbank AD	www.unionbank.bg
18	Municipal Bank AD	www.municipalbank.bg
19	Piraeus Bank Bulgaria AD	www.piraeusbank.bg
20	ProCredit Bank (Bulgaria) AD	www.procreditbank.bg
21	Raiffeisenbank (Bulgaria) EAD	www.raiffeisen.bg
22	Societe Generale Expressbank AD	www.sgeb.bg[15]
23	TBI Bank EAD	www.tbibank.bg[16]
24	Texim Bank AD	www.teximbank.bg
25	Tokuda Bank AD	www.tcebank.com
26	UniCredit Bulbank AD	www.bulbank.bg[17]
27	United Bulgarian Bank AD	www.ubb.bg

Methodology and results from the study

The methodology used in the survey include analyzes of the content of head part of the http response given by the web server. The process was done manually by expert estimation.

[13] Branches of the following banks: Citibank N.A. - Sofia Branch, Regional Investment Bank – Bulgaria Branch and T.C. Ziraat Bankasi - Sofia Branch don't have Bulgarian domains and are not included in the list and in the survey.

[14] In 2009 the domain name was eibank.bg

[15] In 2009 the domain name was sgexpressbank.bg

[16] In 2009 the domain name was westeastbank.bg

[17] In 2009 the domain name was unicreditbulbank.bg

The summarized results of the studied web servers are presented in three tables based on the following key indicators: operating system, type of web server and type of modules installed on the web server. Specific data about which bank what products uses is not provided.

Table 7. Operating System

Operating Systems	Count		Distribution / Type	Count	
	2009	2012		2009	2012
UNIX/GNU LINUX	11	6	AIX	-	1
			CentOS	2	1
			Debian	3	2
			Ubuntu	-	1
			UNIX	3	1
			Red Hat/Fedora	2	-
			FreeBSD	1	-
Windows	11	11			
undetermined	5	10			

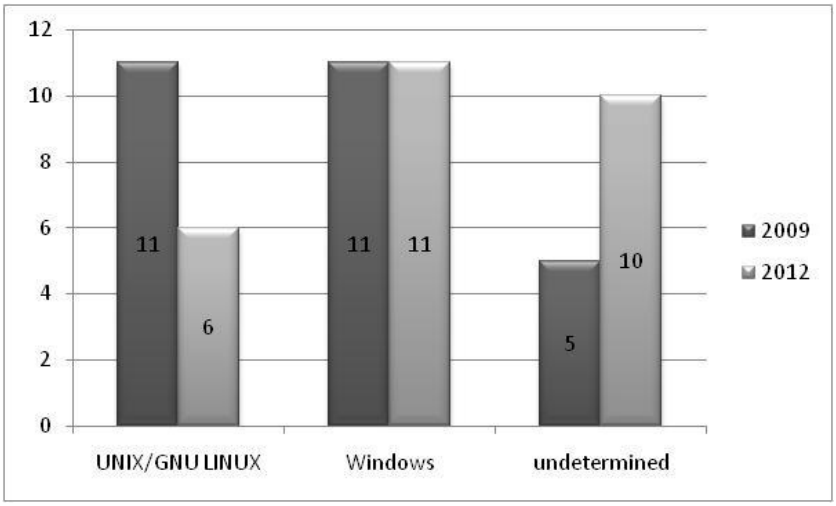


Fig. 5. Relative distribution by Operating Systems

Presented in Table 2 data show that the share of the major operating systems - Windows and UNIX/GNU Linux in 2012 hardly can be determined with an acceptable degree of accuracy due to the high percentage of undetermined operating systems – near 37% in 2012 against 19% in 2009.

When information for the operating system is given, in both periods more web servers use Windows, rather than UNIX / GNU Linux, and the dominance of Windows is over 50% in 2012.

Among the distributions of GNU Linux the most used is Debian, which can be explained by the good support that is available as software packages.

Table 8. Type and version of Web Server

Web Server	Count		Version	Count	
	2009	2012		2009	2012
Apache	15	15	Apache 1.3.x	2	1
			Apache 2.2.x	10	5
			Apache undetermined	3	9
IIS	10	9	IIS/5.0	1	-
			IIS/6.0	9	8
			IIS/7.5	-	1
Lotus-Domino	1	-			
nginx	-	1			
undetermined	1	2			

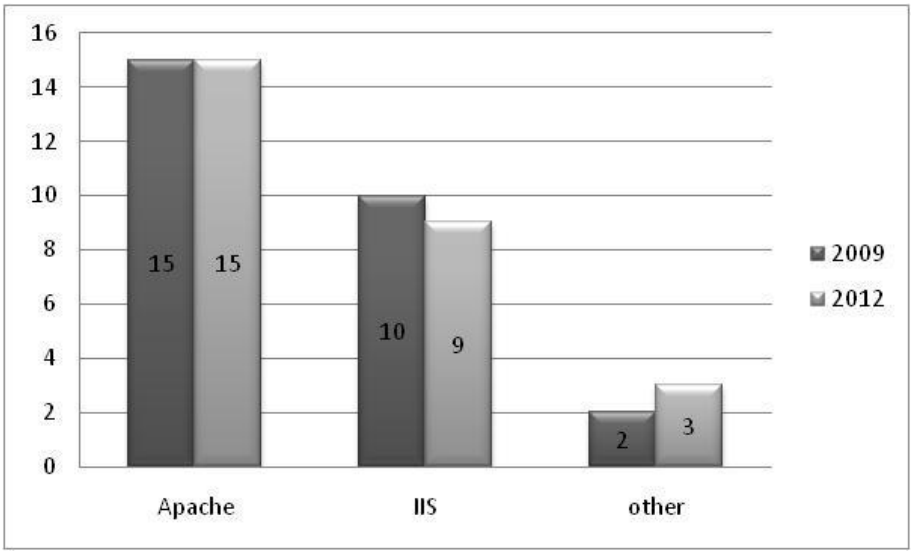


Fig. 6. Relative distribution by web server type

Data in Table 3 shows that in both periods is preferred to use web server Apache - more than half of web servers use it - 56% in 2009 and 2012. Except when do not specified the

version, widely is used a 2.2.x version and only one Bulgarian bank is still using relatively old and dropped out of general use 1.3.x version of Apache.

IIS is the second most popular, but its share is certainly not greater than 40%, even assuming that undetermined web servers are IIS.

One bank in 2012 has starting using web server nginx. In many cases it is used as a reverse proxy server in front of the real web server to dynamically reallocate internally traffic between different web servers for static and dynamically generated content. Nginx is also able to withstand different "Deny of Service" - attacks, and can support a large number of open network connections with relatively used little system resources.

One of the biggest dangers for web servers is the execution of foreign code. IIS6.0 has significant improvements in security over previous versions [2], but there are still a number of vulnerabilities [3, 4]. For IIS7.5 version number of reports of security issues is relatively small.

Table 9. Modules installed on the web server

Module	Count		Version	Count	
	2009	2012		2009	2012
ASP.NET	10	10	ASP.NET/1.1.4322	3	-
			ASP.NET/2.0.50727	5	8
			ASP.NET/4.0.30319	-	1
			ASP.NET undetermined	2	1
PHP	12	12	PHP/4.x	4	-
			PHP/5.x	8	7
			PHP undetermined	-	5
no provided information	5	5			

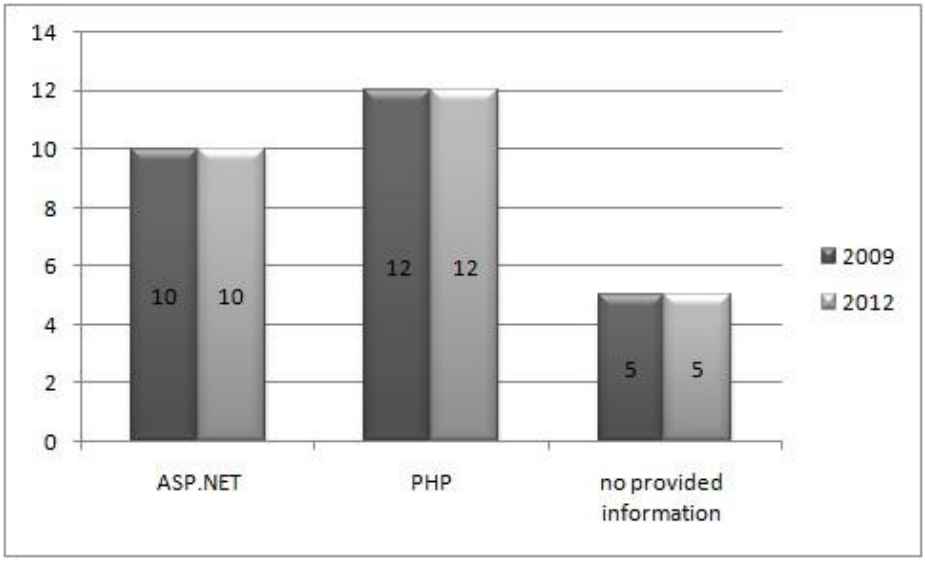


Fig. 7. Relative distribution by installed modules

Data in Table 4 shows that more web servers of Bulgarian banks use PHP modules than ASP.NET modules and this tendency is stable for now.

Conclusion

The reason why we study banks is simple - highly qualified personnel is employed, so decision making is based on the best practices to minimize risks; no lack of funding, so there are opportunities to use any software.

We believe that the web server is more than a key component of the banks IT infrastructure. It has become an integral part of every aspect of bank's online business, so security and performance can't be sacrificed for any reason.

References

1. Bulgarian National Bank, LIST OF LICENSED BANKS AND FOREIGN BANKS' BRANCHES IN THE REPUBLIC OF BULGARIA 03.2012, <http://bnb.bg/BankSupervision/BSCreditInstitution/BSCIRegistrers/index.htm>
2. Belani R., Muckin M. IIS 6.0 Security, 02.11.2010 <http://www.symantec.com/connect/articles/iis-60-security>
3. Microsoft Security Bulletin MS08-006, Vulnerability in Internet Information Services Could Allow Remote Code Execution (942830), 12.02.2008 <http://www.microsoft.com/technet/security/Bulletin/MS08-006.mspx>
4. Microsoft Security Bulletin MS09-020, Vulnerabilities in Internet Information Services (IIS) Could Allow Elevation of Privilege (970483), 09.06.2009, <http://www.microsoft.com/technet/security/Bulletin/MS09-020.mspx>

Evolution of Relational Data Model

Vladimir Dimitrov,

University of Sofia, 5 Bourchier Blvd, Sofia 1164, Bulgaria
cht@fmi.uni-sofia.bg

Abstract. Relational Data Model is frequently accepted as revolution in data modeling. The aim of this paper is to show that it is not true. Relational Data Model is a result of natural evolution based on previous experience and investigations on data modeling.

Keywords: data base, relational data model, data modeling.

1 File Organization

For the first time, relational model of data was published in [1]. Codd intention was to solve some problems with currently used data models. It is better to take an excursion in the past preceding relational model.

In the last 70-ies, first Data Base Management Systems (DBMS) had emerged and many new problems arose in the area of data modeling. DBMS are result of many years, research and investigation on data representation and storage in computing systems.

The most important invention of these days was magnetic disks as primary media for secondary storage. This invention traced then the development of 'file' concept. Initially, the programmer was responsible data placement on the disk sectors and their retrieval. Latter on this responsibility was attached to the operating system. At first, this functionality was implemented as utility programs for file cataloging and disk space management, but then it became part of the operating system. Such an example is the leading these days operating system IBM DOS/360 (Disk Operating System for System 360). For more details on IBM DOS/360 refer [2]. Now the programmer is able to work with files directly supported by the operating system that got responsibility to manage the files: catalog files, pack/unpack records in/from blocks, and allocate/release blocks. Files of records were a new level of abstraction based on the underlying hardware system. Here is the place to mention, that files of bytes (as used in programming languages like C) are abstraction too on the underlying hardware system, but this concept is used to support different kinds of information media. A file of bytes could be placed on magnetic disk, magnetic tape, punched tape, stack of punched cards and so other media. File of bytes is naturally sequentially accessed because all kind of media support sequential read/write. Files of records are naturally randomly accessed. This does not means that file of records could not be sequentially accessed. Files of records with random access are an abstraction of files of sectors on disks. Direct access to records could be supported on magnetic tapes, but it is not efficient.

In random access files, records are ordered in sequence and every record is accessed by its number. This is an abstraction of cataloged file as chain of blocks. The operating system manages a catalog with information about every file and its chain of blocks. It is responsible for consistency of data in the file, i.e. a block is not allocated to two files and data could not

be corrupted by elementary errors as it has been earlier. The operating system transparently allocates and releases blocks to/from files. The blocks in these files are accessed directly by their numbers. A block is a sequence of several sectors. Sector is the elementary read/write unit in magnetic disks. Block is used as a concept in the operating systems for optimization purposes.

Records in files of records are accessed directly by their numbers. This is as in files of blocks and it is the most important concept of files of records. There are some problems related with the files of varying length records. These problems are related to record modification. When a record is updated and its new length is greater than the old length, where the new record could be placed? This new record could not be placed on the old one.

The above mentioned problem could be solved with a new abstraction. Instead identifying records by their sequence numbers a key could be used for this purpose. In this case, random access to the files is based on the keys that uniquely identify records in the file. In such a way records are unpinned from their positioning in the file. With concept of key many file organizations have been developed like indexed-sequential files, B-trees, hash files etc. Even some of them became a language construct in some of the early high level programming languages lake PL/1 and Cobol [3,4].

2 Database Management Systems

The next level of abstraction is based on the file organization. When data is stored in some file organization, this means that some concrete record ordering based on a key is applied. When the file organization for the date is applied, then another ordering by key is applied. This means that an application program that uses some key ordering would fail if it is changed. This is a problem.

First DBMS were based on hierarchy and network organizations of data such systems are IBM IMS [5] and IDMS [6]. They continued to support the concept of record ordering. They offered many utilities for efficient access to records, but their concepts remained in the area of engineering solutions based on file organization. In these DBMS, navigation is based on record ordering and pointers. These DBMS are evolutionary step because their data models do not prescribe specific implementation how this is done in file organizations.

At first glance, relational model do not participate in this evolution. It is based on mathematical theory of sets; it is not further abstraction of some elements of file organization. But relational model is a jump in the evolution of data representation. It is further abstraction of data organization and navigation.

Codd had been motivated to develop a new model of data independent of application programs and way of data access. Now, data access efficiency became DBMS responsibility. Application program, on the other hand, has more simplified and more abstract interface to the date through DBMS. Relational model represents data in their natural structuration without any elements of their machine implementation, how this has been done in previous models of data. Navigation in relational model is available through high level query language. In such a way, application program only declares needed navigation functionality and DBMS implements it.

Relational model is evolution (more precisely jump in the evolution), because Codd intention was to develop new model of data to overcome pitfalls of currently available models of data. The main target was to represent the data to applications in an independent

way of its storage. There three dependencies that relational model had to fight: ordering, indexing and access path dependencies. In the next, a brief discussion on their nature is given.

Ordering dependence. Elements of data could be stored in the data in different ways: in some cases, ordering is not important; in other cases, the element could participate in only one ordering; or the element could participate in several orderings. If an application uses some ordering of data and this ordering is changed, then the application will fail. The pre-relational models do not differentiate representation ordering from storage ordering of data.

Indexing dependence. Creation of indexes on the data permits to accelerate access to data, but it decreases data modifications: update, insert and delete operations. In environments with changeable characteristics of work with data, it is hazardous to use indexes. The problem is the impact of dynamic index creation/removal on applications whose access to the data uses indexes. DBMS administration of index management can solve this problem – not application programmer.

Dependence of access paths. In pre-relational models of data, navigation must follow the hierarchy or graph structuring of data. This means that application programs are bound to the concrete structuring of data. So, every change in data structure could result in some application program fail.

The intention of relational model is to solve above mentioned problems – it is not abstract mathematical construction.

3 Relational Model of Data

Relational model is based on the concept 'relation' from mathematical theory of sets. Relation is a subset of Cartesian product of several domains. Relation is represented as an array with columns named by the domains participating in the Cartesian product. This array has as many rows as tuples has the relation. Every column of this array contains values only from the domain that names it. Every row corresponds to a tuple of the represented relation.

Codd names columns by the domains. This is problem when a domain is used more than one time in the Cartesian product. To resolve this ambiguity, Codd adds a role to distinguish usage of such domains. Using domains (and roles), Codd tries to give way for specification of data semantic.

Formally, relational model is presented in [7], where this initial view on the relations is discussed in details.

Initial implementations of relational model showed deviations from presentation in [1]. For example, semantics of domains is the corner stone of semantics of relations, but these first DBMS do not have mechanisms for complex domain management, instead they have as domains the sets of types of conventional programming languages. The last ones are taken from mathematics and they are very abstract to contain any suitable semantics for data modeling of the real world. That is why Codd, later on, tried several times to redefine relational model to capture more meaning as in [8].

In relational model, data structures are independent of ordering. Relation is a set of tuples and navigation is independent of their ordering. There is a freedom to represent the same relation with different file organizations.

One of the questions that arise is why Codd has used concept of relation as fundamental structure in the model instead of more generalized concept of set? In reality, the relation is a

specialized kind of set. In many models of data basic element is the set. Codd's selection has deep roots in theory and practice. First, relation is an abstraction of data storage on disks. There, data are stored in atomic units – sectors with fixed length. The sector is the smallest unit for read/write operations on the disk. For optimization purpose could be used block that are multiple sequential sectors, but this does not change the situation. Relation could be viewed as a machine representation of data on the disk where relation corresponds to disk and tuples – to sectors.

Every domain has some representation of its values. This representation could be fixed or varying one. Tuples could be of fixed or varying length depending on representation of relation domains values. Varying length tuples are represented with fixed length blocks. Efficiency of access to tuples depends of efficiency of access to fixed length blocks. Tuples are abstraction of the records from underlying file system. Tuples do not put any restrictions on the record representation in the different file organizations. This means that relation representation could be freely changed from one file organization to another.

Relational model is in reality abstraction of magnetic disk structure. From the time of relational model emergence till now, there is no basic change in the concept of secondary storage media – magnetic disks. That is why relational model is still alive.

There are some alternative implementations of relational model that are not based the abstraction tuple-record and relation-file. For example, some of these implementations try to represent relation as relationship among its domains; the values are stored in domain structure; and the relation structure does not contain domain values, but only pointers to them. This approach does not follow the natural evolution of data representation and has not received industry support.

A variation on relational model is binary model of data [10], where all relations are binary ones. This approach again is deviation from the evolution and has not received industry support.

So, the relation, how was introduced from the beginning, was an abstraction of file organization with random access by key. This relation has one dimension of freedom – independence of tuple ordering, but still there is a problem with the fixed domain ordering. This is a problem from user's point of view – it is not implementation problem. The user has to know relation name, the names of its domains (and roles for duplicated domains) and domains ordering. Codd solved this problem introducing 'relationships'. This is a relation in which domain ordering is not important. Relationship represents all relations that could be produced by all permutations of its domains. Now, the user has to know relationship name and its domain names. But even this information could be derived with some DBMS utilities. Relationship is closer to nowadays term 'table'. More detailed discussion and formal specification can be found in [11]. In this specification the database is a set of relationships implemented with a database of relations how the things happen to be in the implementations. So, from users point of view DBMS manages relationships, but from implementation point of view – relations.

In this presentation of relational model, there are no indexes, pointers or access paths. Instead, there is a high level query language – relational algebra. DBMS must implement this query language to support relational model. Relational algebra is an abstraction of navigation without any access paths, indexes and pointers. It is based on set operations with some extensions. This is a declarative query language – not navigational one. The application program has to declare its data needs without specifying any access paths or indexes. DBMS job is to retrieve and return these data. Access paths and indexes are part of DBMS

implementation – not of relational model. The application program is not bounded with any details on the navigation. Relational model does not depend on access paths and indexes. Relational algebra on relations is formally specified in [12]. The operations in this specification are derived from [1].

4 Conclusion

Relational model is a result of natural evolution in data modeling. It is a jump in the evolution that has deep roots in previous data models and underlying hardware structure. That is why relational model is still on the stage and every Computer science (CS) and Information system (IS) curriculum include at least one database management course based on the relational model [13].

Acknowledgment. This paper is supported by Sofia University "St. Kliment Ohridski" SRF under Contract 134/2012.

References

1. Codd, E.F.: A Relational Model of Data for Large Shared Data Banks. CACM vol. 18, no 6, 377--387 (1970)
2. DOS/360 and successors: http://en.wikipedia.org/wiki/DOS/360_and_successors
3. PL/I: http://en.wikipedia.org/wiki/PL/I
4. COBOL:, http://en.wikipedia.org/wiki/COBOL
5. IBM Information Management System: http://en.wikipedia.org/wiki/IBM_Information_Management_System
6. IDMS: http://en.wikipedia.org/wiki/IDMS
7. Dimitrov, V.: Formal Specification of Relational Model of Data in Z-Notation, Proc. of XXXIX Conf. of UMB, 178-183 (2010)
8. Codd, E.F.: Extending the Database Relational Model to Capture More Meaning, ACM Transactions on Database Systems, Vol. 4, No. 4, 397—434 (1979)
9. Seshadri, P., Livny, M., Ramakrishnan, R.: SEQ: Design and Implementation of Sequence Database System, Proc. of the 22nd Int. Conf. on Very Large Data Bases (VLDB '96), Mumbay, India, 99—110(1996)
10. Abrial, J.R.: Data Semantics, Proc. of IFIP Conf. on Data Base Management, North-Holland, 1—59 (1974)
11. Dimitrov, V.: "Relationship" Specified in Z-Notation, Physics of Elementary Particles and Atomic Nuclei, Letters, Vol. 8, No. 4(167), 655—663 (2011)
12. Dimitrov, V.: Formal Specification of Relational Model with Relational Algebra Operations, Proc. of Fifth Int. Conf. ISGT, 25-28 May 2011, Sofia, pp. 57-74.
13. Kaloyanova K., Using the Project Approach in IS Education, Proc. of the 8th International Conference Computer Science and Education in Computer Science (CSECS), Boston- New York, 5-9 July, 2012, pp 36-40

Finite State Automata Semantics in Communicating Sequential Processes

Vladimir Dimitrov,

University of Sofia, 5 Bourchier Blvd, Sofia 1164, Bulgaria
cht@fmi.uni-sofia.bg

Abstract. Finite state automata have been used for specification of business processes. Communication sequential processes are more powerful tool used in modern business process tools. This paper is devoted on formal specification of finite state automata as communicating sequential processes.

Keywords: finite state automata, communication sequential processes, formal specification.

1 Motivation

Traditionally, distributed systems and protocols are described with finite state automata (finite state machines). As result of that, many tools based on finite state automata have been developed and used. Such an example is business state machines used in IBM WebSphere Integration Developer [1].

Later on, other more powerful mathematical tools for specification and analyses of distributed systems have been developed, such as Petri nets [2], CSP [3], and so on. Modern tools and notations for specification, development and implementation of distributed systems are based on them. For example, Petri nets concepts are broadly used for specification of business processes in notations like UML (activity diagrams) [4], WS-BPEL [5], BPMN [6], etc. The newer mathematical tools are more powerful that the older ones. For example, Petri nets have more expressive power than finite state automata, but are less expressive than CSP. Our intention, here in this paper, is not to compare them.

In commercial tools, that use finite state automata, as a base for business process specification, the problem is the need to convert older specifications into new one without losing the semantics.

Newly developed tools are usually based on Petri nets or CSP. They are more powerful in specification and analyses, but they have to support continuity with the older developments. Such an example is IBM WebSphere Integration Developer that nowadays is based on WS-BPEL, but has to support backward compatibility with the business state machines.

Intention of this paper is formally to specify finite state automata in CSP. Finite state automata semantics is clear, but there are needs for conversion of business processes specified in them to new form without losing the semantics.

2 Definitions

There are many kinds of finite state automata: deterministic, non-deterministic, Mealy machines, Moor machines etc. Some extensions like Turing machine get outside the expressive power of finite state automata, but they are not subject of this paper.

We will use the next definition of finite state automata: Finite state automata A with alphabet V is the 5-tuple A= <K, V, δ, q0, F>, where:

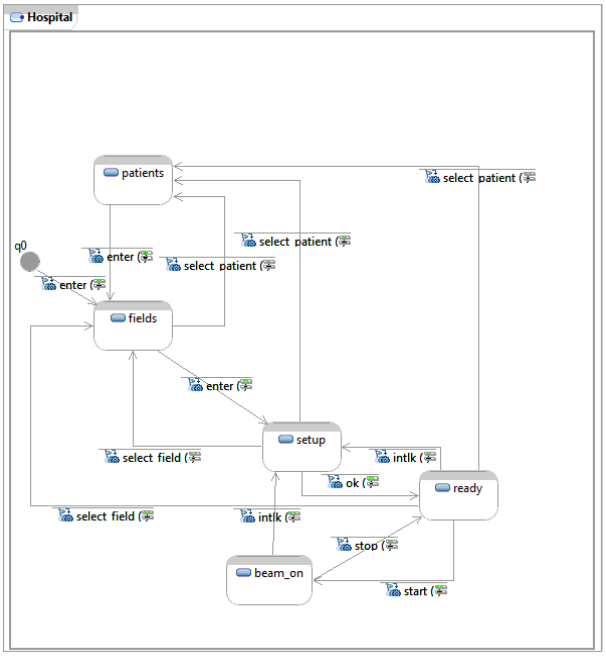
- K is non empty finite set of automata states;
- V is non empty finite set of input symbols - the alphabet;
- δ is transition function with domain K x V and range K;
- q0 ∈ K is the initial state;
- F ⊆ K is the set of final states.

It is possible F to be empty, in this case the machine is executed forever or to stop not in final state. When automata stop in finite state it has finished normally its work, but if it stops in non final state – this means that machine is broken in some way.

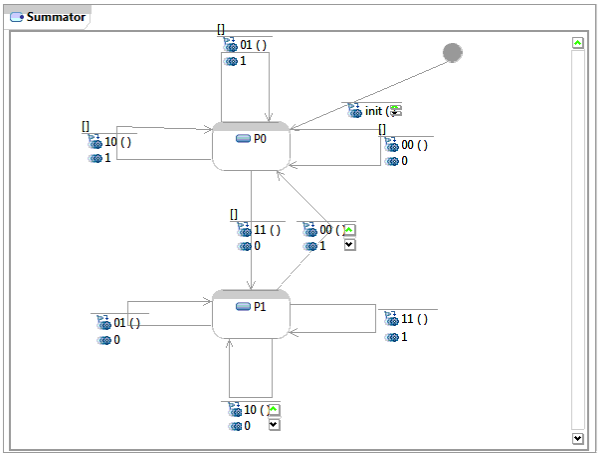
A finite state automata is deterministic if its transition function is defined in every state for all input symbols, i.e. domain δ is equal to K x V. If the finite state automata has at least one state for which the transition function is not defined for all input symbols, then this automata is non deterministic. From the theory, we know that every non deterministic machine can be modeled deterministic one. This means that they are equivalent in expressiveness.

In some definitions of finite state automata output alphabet O and output function ω are included. When a transition is executed, it is possible to be generated some output. Domain of ω is subset of the domain of δ, but its range is O.

Some examples follow. First example:



There are no final states (F = ∅). This machine is non-deterministic with:
K = {q0, patients, fields, setup, ready, beam_on}
V = {select_patient, select_field, enter, ok, start, stop, intlk}
δ = {(q0, enter) ↦ fields, (patients, enter) ↦ fields,
(fields, select_patient) ↦ patients, (fields, enter) ↦ setup,
(setup, select_patient) ↦ patients, (setup, select_field) ↦ fields,
(setup, ok) ↦ ready, (ready, select_patient) ↦ patients,
(ready, select_field) ↦ fields, (ready, start) ↦ beam_on,
(ready, intlk) ↦ setup, (beam_on, stop) ↦ ready, (beam_on, intlk) ↦ setup}
In the second example, the machine has output:



where:
K = {q0, P0, P1}
V = {init, 00, 01, 10, 11}
δ = {(q0, init) ↦ P0, (P0, 00) ↦ P0, (P0, 01) ↦ P0, (P0, 10) ↦ P0, (P0, 11) ↦ P1,
(P1, 00) ↦ P0, (P1, 01) ↦ P1, (P1, 10)↦ P1, (P1, 11) ↦ P1}
O = {NULL, 0, 1}
ω = {(q0, init) ↦ NULL, (P0, 00) ↦ 0, (P0, 01) ↦ 1, (P0, 10) ↦ 1, (P0, 11) ↦ 0,
(P1, 00) ↦ 1, (P1, 01) ↦ 0, (P1, 10) ↦ 0, (P1, 11) ↦ 1}

3 Specification in Z-notation

Here, we will be more strict specifying finite state automata in Z-notation [7].
Basic sets are:

[*STATES*, *INPUTS*, *OUTPUTS*]

where STATES is non empty final sets of automata states, INPUTS is the set of input symbols (events), OUTPUTS is the set of all output symbols.

$q0$: *STATES*
NULL: *OUTPUTS*
FINALS: $\mathbb{F}$ *STATES*

$STATES \neq \emptyset \land (\exists n: \mathbb{N} \bullet \#STATES \leq n) \land$
$INPUTS \neq \emptyset \land q0 \notin FINALS$

where q0 is the initial state, NULL is a special output symbol – nothing is outputted, FINALS is possible empty subset of STATES – final states.

FSM

transition: $STATES \times INPUTS \nrightarrow STATES$
output: $STATES \times INPUTS \nrightarrow OUTPUTS$
current: *STATES*

---

dom *output* = *dom transition* $\land$
$q0 \in$ dom (dom *transition*) $\land$
$FINALS \cap$ dom (dom *transition*) $= \emptyset \land$
$FINALS \subseteq$ ran *transition* $\land$
$STATES \setminus \{q0\}$ = ran *transition* $\land$
$STATES \setminus FINALS$ = dom (dom *transition*)

Finite states machine consists of transition function, output function and current state. Transition and output functions have the same domain Cartesian product of STATES and INPUTS. The initial state q0 is part of the domain of transition function. FINALS states have only input arcs, but no output arcs. Only q0 has no input arcs. All states, except final ones, have to have input and output arcs.

FSMInit

FSM

---

current = $q0$

Finite states automate initially starts with current state q0.

Execute

ΔFSM
$i?: INPUTS$
$o!: OUTPUTS$

$(current, i?) \in \text{dom } transition \land$
$current' = transition(current, i?) \land o! = output(current, i?) \land$
$transition' = transition \land output' = output$

Execution of finite state automata consists of application of functions transition and output to the input and the current state. Current state is modified in successful transition. Here is not defined what the automata will do is an unexpected input in this state is accepted. There are to possible actions: the first one is not to react and the state to remain the same; the second is to indicate an error. What approach will be used depend of automata nature. If the machine is grammar recognition it must react with error. If the automata is not of that that type it is possible simply to ignore the input and to remain in the same state. In any case, it is possible the automata to be represented with deterministic one and then transition function will be total and no such a problem will arise.

Finally, we will do some comments about finite state automata and business processes. There are two kinds of processes: such that are executed one time and finish and such that are started and then execute endless. In any case, they have to be initialized in some way that is why initial state is obligatory. But in the first case final states are obligatory. The business process cannot stop in another state (except in final states) – this is design error and has to be checked.

Now, let's see examples. The first example:

$STATES ::= q0 \mid patients \mid fields \mid setup \mid ready \mid beam\_on$
$INPUTS ::= select\_patient \mid select\_field \mid enter \mid ok \mid start \mid stop \mid intlk$

$FINALS: \mathbb{F}\ STATES$

$STATES \neq \emptyset \land (\exists n: \mathbb{N} \bullet \#STATES \leq n) \land$
$INPUTS \neq \emptyset \land q0 \notin FINALS \land FINALS = \emptyset$

FSM

$transition : STATES \times INPUTS \nrightarrow STATES$

$current : STATES$

$transition = \{(q0, enter) \mapsto fields, (patients, enter) \mapsto fields,$
$(fields, select\_patient) \mapsto patients, (fields, enter) \mapsto setup,$
$(setup, select\_patient) \mapsto patients, (setup, select\_field) \mapsto fields, (setup, ok) \mapsto ready,$
$(ready, select\_patient) \mapsto patients, (ready, select\_field) \mapsto fields,$
$(ready, intlk) \mapsto setup, (ready, start) \mapsto beam\_on,$
$(beam\_on, stop) \mapsto ready, (beam\_on, intlk) \mapsto setup \} \land$
$q0 \in \mathrm{dom}\ (\mathrm{dom}\ transition) \land$
$FINALS \cap \mathrm{dom}\ (\mathrm{dom}\ transition) = \emptyset \land$
$FINALS \subseteq \mathrm{ran}\ transition \land$
$STATES \setminus \{q0\} = \mathrm{ran}\ transition \land$
$STATES \setminus FINALS = \mathrm{dom}\ (\mathrm{dom}\ transition)$

FSMInit

FSM

$current = q0$

Execute

ΔFSM

$i? : INPUTS$

$(current, i?) \in \mathrm{dom}\ transition \land$
$current' = transition(current, i?) \land$
$transition' = transition$

OUTPUTS and output function are eliminated in this specification, but STATES, INPUTS and transition are in full details. No output parameter during execution.

The second example is:

$STATES ::= q0 \mid P0 \mid P1$

$INPUTS ::= init \mid i00 \mid i01 \mid i10 \mid i11$

$OUTPUTS ::= NULL \mid o0 \mid o1$

$$FINALS : \mathbb{F}\ STATES$$

$$STATES \neq \emptyset \land (\exists n : \mathbb{N} \bullet \#STATES \leq n) \land$$
$$INPUTS \neq \emptyset \land q0 \notin FINALS \land FINALS = \emptyset$$

FSM

$$transition : STATES \times INPUTS \nrightarrow STATES$$
$$output : STATES \times INPUTS \nrightarrow OUTPUTS$$
$$current : STATES$$

$$transition = \{(q0, init) \mapsto P0,$$
$$(P0, i00) \mapsto P0, (P0, i01) \mapsto P0, (P0, i10) \mapsto P0, (P0, i11) \mapsto P1,$$
$$(P1, i00) \mapsto P0, (P1, i01) \mapsto P1, (P1, i10) \mapsto P1, (P1, i11) \mapsto P1\} \land$$
$$output = \{(q0, init) \mapsto NULL,$$
$$(P0, i00) \mapsto o0, (P0, i01) \mapsto o1, (P0, i10) \mapsto o1, (P0, i11) \mapsto o1,$$
$$(P1, i00) \mapsto o1, (P1, i01) \mapsto o0, (P1, i10) \mapsto o0, (P1, i11) \mapsto o1\} \land$$
$$\text{dom } output = dom\ transition \land$$
$$q0 \in \text{dom (dom } transition) \land$$
$$FINALS \cap \text{dom (dom } transition) = \emptyset \land$$
$$FINALS \subseteq \text{ran } transition \land$$
$$STATES \setminus \{q0\} = \text{ran } transition \land$$
$$STATES \setminus FINALS = \text{dom (dom } transition)$$

FSMInit

$$FSM$$

$$current = q0$$

Execute

$$\Delta FSM$$
$$i? : INPUTS$$
$$o! : OUTPUTS$$

$$(current, i?) \in \text{dom } transition \land$$
$$current' = transition(current, i?) \land o! = output(current, i?) \land$$
$$transition' = transition \land output' = output$$

Here, only final states are not defined.

4 Into the CSP

Let finite state machine is defined as follows:

- STATES = $\{q_0, q_1, \dots, q_n\}$ is the set of states;
- INPUTS = $\{i_1, i_2, \dots, i_m\}$ is the set of input symbols;
- OUTPUTS = $\{o_1, o_2, \dots, o_p\}$ is the set of output symbols;
- FINALS = $\{f_1, f_2, \dots, f_q\}$ is the set of final states.

The communicating sequential process P modeling finite state machine is represented as a choice:

P = {x: B -> P(i)}

where B is the set of indexes of the states:

B = 0..n

and then:

P = {i: 0..n -> P(i)}

This process communicates with the environment via two channels *in* and *out*. The channels and process alphabets are:

$\alpha(in) = \{ i_1, i_2, \dots, i_m\}$

$\alpha(out) = \{o_1, o_2, \dots, o_p\}$

$\alpha(P) = \alpha(in) \cup \alpha(out)$

Every expression P(i) is represented by a process modeling finite state automata behavior in state q_i:

P(i) = Pi, i = 0, … , n

Let's see now what is Pi. If q_i is a final state then:

Pi = SKIP

If q_i is not a final state then the transition function is defined for q_i and some subset of input events $\{i_{i1}, i_{i2}, \dots, i_{is}\}$. Let these transitions be:

$(q_i, i_{ij}) \mapsto q_{ij}$ for j = 1, … , s

The subexpression for this transition is:

$in?i_{ij}$ -> Pij for j = 1, … , s

If the output is defined for this transition, i.e.:

$(qi, i_{ij}) \mapsto o_{ij}$

then this subexpression will be:

$in?i_{ij}$ -> $out!o_{ij}$ -> Pij

The whole Pi is:

Pi = in? i_{i1} -> $out!o_{i1}$ -> Pi1 | $in?i_{i2}$ -> $out!o_{i2}$ -> Pi2 | … | $in?i_{is}$ -> $out!o_{is}$ ->Pis

Note: output communications are not defined for all transitions.

Now, let's see how this looks for the examples.

Example 1: Hospital

Hospital = {s: {q0, patients, fields, setup, ready, beam_on } -> Q(s)}

Qstart = in?enter -> Qfields

Qpatients = in?enter -> Qfields

Qfields = in?select_patient -> Qpatients | in?enter -> Qsetup

```
Qsetup = in?select_patient -> Qpatients |
        in?select_field -> Qfileds | in?on -> Qready
Qready = in?select_patient -> Qpatients |
        in?select_field -> Qfileds | in?intkl -> Qsetup
        | in?start -> Qbeam_on
Qbeam_on = in?intkl -> Qsetup | in?stop -> Qready
```

Example 2: Summator

```
Summator = {s: {q0, P0, P1} -> P(s)}
Pq0 = in?init -> Pp0
Pp0 = in?00 -> out!0 -> Pp0 | in?01 -> out!1 -> Pp0 | in?10 -> out!1 -> Pp0 |
        in?11 -> out!0 -> Pp1
Pp1 = in?00 -> out!1 -> Pp0 | in?01 -> out!0 -> Pp1 | in?10 -> out!0 -> Pp1 |
        in?11 -> out!1 -> Pp1
```

Finally, some comments about this CSP models. In both examples, finite state automata are not recognition ones, because if some event that is not accepted in given state, is send via the channel the process will not react on it. For recognition purposes deterministic finite state machine has to be used or some additional processing of errors has to be added.

5 Implementation in PAT 3

Process Analysis Toolkit [8] is an enhanced simulator, model checker and refinement checker for concurrent and real-time systems. It implements a version of CSP. Let's see our examples implemented in PAT 3.

Example 1: Hospital

```
enum {q0, patients, fields, setup, ready, beam_on};
enum {select_patient, select_field, enter, ok, start, stop, intlk, end};
#define error -1;
channel in 0;
channel out 0;
#alphabet Q{in.enter, in.select_patient, in.select_field, in.ok, in.intlk, in.start, in.stop,
in.end};
Q(state) = case {
        state == q0: in?enter -> Q(fields) [] in?end -> Skip
        state == patients: in?enter -> Q(fields) [] in?end -> Skip
        state == fields: in?select_patient -> Q(patients) [] in?enter -> Q(setup)
                [] in?end -> Skip
        state == setup: in?select_patient -> Q(patients) [] in?select_field -> Q(fields)
                [] in?ok -> Q(ready) [] in?end -> Skip
        state == ready: in?select_patient -> Q(patients) [] in?select_field -> Q(fields)
                [] in?intlk -> Q(setup) [] in?start -> Q(beam_on) [] in?end -> Skip
```

```
        state == beam_on: in?intlk -> Q(setup) [] in?stop -> Q(ready)
                [] in?end -> Skip
        default: out!error -> Stop
};
System() = in!enter -> in!enter -> in!ok -> in!start -> in!end -> Skip ||| Q(q0);
#assert System() deadlockfree;
#assert System() deterministic;
```

Here, events and states are defined as constants (named numbers) with enum. An event error is defined, because there is no other way of control on process parameters. If an error parameter is accepted by the process, then on out channel error event is sent. Input and output channels have to be defined not buffered. The alphabet of the process is restricted to the given one.

The main difference is in the implementation of the process; instead for every choice alternative to be delivered as different process, process expressions are included directly in the choice operator. The choice is CSP is simply an operator defined on a set of events, but here in this implementation choice can be defined on process parameters. This idea processes to have parameters, is used in some versions of CSP, but not in the original representation.

Finally, the system can be checked for many properties like deadlock free, determinism etc. These checks are put at the end of the specification and can be verified, but only processes without parameters can be verified, that is why such a process communicating with the machine is defined and checked.

Example 2: Summator

```
enum {q0, P0, P1};
enum {initialize, i00, i01, i10, i11, end};
enum {o0, o1, error};
channel in 0;
channel out 0;
#alphabet P{in.initialize, in.i00, in.i01, in.i10, in.i11, o.0, o.1};
P(state) = case {
        state == q0: in?initialize -> P(P0) [] in?end -> Skip
        state == P0: in?i00 -> out!o0 -> P(P0) [] in?i01 -> out!o1 -> P(P0)
                [] in?i10 -> out!1 -> P(P0)
                [] in?i11 -> out!o0 -> P(P1) [] in?end -> Skip
        state == P1: in?i00 -> out!o1 -> P(P0) [] in?i01 -> out!o0 -> P(P1)
                [] in?i10 -> out!0 -> P(P1)
                [] in?i11 -> out!o1 -> P(P1) [] in?end -> Skip
        default: out!error -> Stop
};
System() = in!initialize -> in!i00 -> out?x -> in!end -> Skip ||| P(q0);
#assert System() deadlockfree;
#assert System() deterministic;
```

In this example, output function is included. One more addition is that end event is added to stop the machine in every state.

This process can be implemented with a global variable instead of process parameters, like that:

```
enum {q0, P0, P1};
enum {initialize, i00, i01, i10, i11, end};
enum {o0, o1, error};
channel in 0;
channel out 0;
var state = q0;
#alphabet P{in.initialize, in.i00, in.i01, in.i10, in.i11, o.0, o.1};
P() = case {
        state == q0: in?initialize -> {state = P0} -> P [] in?end -> Skip
        state == P0: in?i00 -> out!o0 -> {state = P0} -> P
                [] in?i01 -> out!o1 -> {state = P0} -> P
                [] in?i10 -> out!1 -> {state = P0} -> P
                [] in?i11 -> out!o0 -> {state = P1} -> P
                [] in?end -> Skip
        state == P1: in?i00 -> out!o1 -> {state = P0} -> P
                [] in?i01 -> out!o0 -> {state = P1} -> P
                [] in?i10 -> out!0 -> {state = P1} -> P
                [] in?i11 -> out!o1 -> {state = P1} -> P [] in?end -> Skip
        default: out!error -> Stop
};
System() = in!initialize -> in!i00 -> out?x -> in!end -> Skip ||| P();
#assert P() deadlockfree;
#assert P() deterministic;
```

But this implementation is not so clear and diverges from CSP. In above example, some properties are impossible to be checked, because the verifier is not sure about the contents of the global variable state. It finds out that the process is deterministic, but thinks that it is not deadlock free. Here, we will stop and will not go in further details what is possible and what is not to do with the tool.

6 Conclusion

Finite state machines successfully could be modeled in CSP. This means that old process specifications can be converted into the CSP without loss of semantics.

Acknowledgment. This paper is supported by ДДВУ02/22/20.10.2011 with National Science Fund.

References

1. IBM WebSphere Process Server and WebSphere Integration Developer, Version: V6.0.2, Business state machines, http://publib.boulder.ibm.com/infocenter/ieduasst/v1r1m0/ index.jsp?topic=/com.ibm.iea.wpi_v6/wpswid/6.0.2/BusinessStateMachine.html

2. Petri net, http://en.wikipedia.org/wiki/Petri_net
3. Communicating sequential processes, http://en.wikipedia.org/wiki/Communicating_sequential_processes
4. UML, http://www.uml.org
5. OASIS Web Services Business Process Execution Language (WSBPEL) TC, http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsbpel
6. BPMN, http://www.bpmn.org/
7. Z-notation, http://en.wikipedia.org/wiki/Z_notation
8. Process Analysis Toolkit, http://www.comp.nus.edu.sg/~pat/10. Abrial, J.R.: Data Semantics, Proc. of IFIP Conf. on Data Base Management, North-Holland, 1—59 (1974)

ASSESSMENT OF RISK OBJECTS

Juliana Karakaneva, Katya Petrova

Abstract. In this paper the authors present a conceptual model that can be used to create a tool for assessing of risk objects. The model combines the advantages of neural networks and multiatribute utility theory. This model provides intelligent control of risk objects. In order to minimize the risk of disclosure or data lost delta learning algorithm is used. The algorithm is based on the least-square error minimization method. The objective with this method is to express the difference of actual and target outputs in terms of the inputs and weights. We start with the least-squared error between the target output and actual output. The learning rate is recomputed at each iteration. If the error is larger than a desired value the calculations are repeated until the desired minimum error is reached. The hybrid model for assessment of risk events, which make an impact on risk objects, is useful for decision makers in order to prognosticate and prevent adverse effects on security.

Keywords: risk assessment, risk management, ISMAUT, Delta Algorithm.

Introduction

"Risk Objects" are features, activities or situations to which harmful consequences are conceptually attached. To become risk objects risk events or incidents occur.

To ensure information security events and weaknesses associated with information, systems are communicated in a manner allowing timely corrective action to be taken. Formal event reporting and escalation procedures should be in place. All employees, contractors and third party users should be made aware of the procedures for reporting the different types of event and weakness that might have an impact on the security of organizational assets.

To prevent the risk events the decision makers have to analyse and evaluate the threats and risks for security. Several approaches are applied in this area, including diagnostic expert system, statistical models and evaluations, event chain methodology, neural network and other artificial intelligence techniques.

This paper proposes a hybrid model for assessment of risk events which make an impact on risk objects. The model combines the advantages of Multi-Attribute Utility Theory (ISMAUT) [1] and neural networks (Delta learning algorithm) [2].

Reporting information security events

Information security events should be reported through appropriate management channels in real time. A formal information security event reporting procedure should be established, together with an incident response and escalation procedure, setting out the action to be taken on receipt of a report of an information security event. A point of contact should be established for the reporting of information security events. It should be ensured

that this point of contact is known throughout the organization, is always available and is able to provide adequate and timely response.

All employees, contractors and users should be report information security events to the appropriate point of contact. The reporting procedures should include [3]:

- suitable feedback processes to ensure that those reporting information security events are notified of results after the issue has been dealt with and closed;
- information security event reporting forms to support the reporting action, and to help the person reporting to remember all necessary actions in case of an information security event;
- the correct behaviour to be undertaken in case of an information security event, i.e. noting all important details (e.g. type of non-compliance or breach, occurring
  malfunction, messages on the screen, strange behaviour) immediately;
- not carrying out any own action, but immediately reporting to the point of contact;
- reference to an established formal disciplinary process for dealing with employees, contractors or users who commit security breaches.

With due care of confidentiality aspects, information security incidents can be used in user awareness training as examples of what could happen, how to respond to such incidents, and how to avoid them in the future. To be able to address information security events and incidents properly it might be necessary to collect evidence as soon as possible after the occurrence. Malfunctions or other anomalous system behavior may be an indicator of a security attack or actual security breach and should therefore always be reported as information security event.

Imprecise Value Functions

There are many planning and decision making situation, which are characterized by multiple conflicting objectives, and a multiattribute value function can represent the preference structure needed to solve the problem.

The vector of weights is defined with a methodology of Imprecisely Specified Multiattribute Utility Theory. For realization of the method the follows components are necessary:

- a list of alternatives,
- a list of attributes,
- a precise calculation of each alternative for each attribute,
- a set of weights that induce relatively participation of each attribute.

The overall weighted average of alternatives is defined as function of the alternatives and the weights. Benefits can be specified as a set of linear inequalities.

Two conditions must be presented for a multiattribute value function to exist. The first condition is a monotonicity condition, which states if the value of one attribute improves while there is no loss in value for other attributes, preference must increase. The second is a continuity condition, which states if a_i f a_j f a_k (means that a_i is preferred to a_j), then there must be a unique point where the decision maker is indifferent between the increase from a_k to a_j and the increase from a_j to a_i [5].

The form of a decomposable value function here is an additive value function. This function is a weighted sum of attribute values, such that the value of an alternative with k attributes is given by:

$$v = \sum_k w_k \, v_k(r_k), \qquad (1)$$

where w_k is the tradeoff weight for attribute k, $v_k(r_k)$ is the value produced by the value function of attribute k and r_k is the level of attribute k. Keeney and Raiffa (1976) describe the process for determining the attribute value functions and tradeoff weights. This form of the multiattribute value function requires that the attributes be mutually, preferentially independent.

Definition[5]. A set of attributes is mutually, preferentially independent if every subset of these attributes is preferentially independent of its complementary set.

It can also be shown that in general if each pair of attributes is preferentially independent of its complement, then (1) is valid (Gorman, 1968). Unlike standard value functions, the additive value function is unique up to a positive linear transformation.

In addition to decomposing a multi-attribute value function, the amount of quantitative analysis required to construct the function can be reduced by using an imprecise value function. ISMAUT (White et al., 1984) creates a partial order based on preference relationships among a finite set of alternatives A. The method uses a weighted sum of attribute values. The value of alternative a_i is given by (1). A preference statement of the form a_i f a_j implies an inequality in the space of possible weights according to the following relation:

$$\sum_k v_k = [v_k(r_{ik}) - v_k(r_{jk})] \geq 0 \qquad (2)$$

According to this interpretation, the statement that a_i is preferred to a_j means that the tradeoff weights are such that the total weighted value of a_i is at least as great as that of a_j.

Furthermore, the sum of weights must unity:

$$\sum_k w_k = 1 \qquad (3)$$

All these inequalities confine the weight space to a subspace, W', that satisfies the inequalities. Thus, from pairwise preference statements, the method determines ranges of attribute weights consistent with decision maker's preferences. The imprecise value function v can order pairs other than those specified: a_i is preferred to a_j if, for every possible vector of weights <$w1$, $w2$, ..., wk> within W', the value of a_i is greater than the value of a_j, i.e.,

$$Min \sum_k [v_k(r_{ik}) - v_k(r_{jk})] \geq 0 \qquad (4)$$

This relation can be tested for every pair of alternatives that the decision maker has not already stated a preference. Thus, the specified preferences create a partial order over all alternatives, and this partial order can identify the nondominated set of alternatives. The following example illustrates how implicit preference relations can be derived.

Delta Learning Algorithm

The Delta learning algorithm is based on the least-square-error minimization method. The objective with this method is to express the difference of the actual and target outputs in terms of the inputs and weight [6]. Figure 1 presents this method.

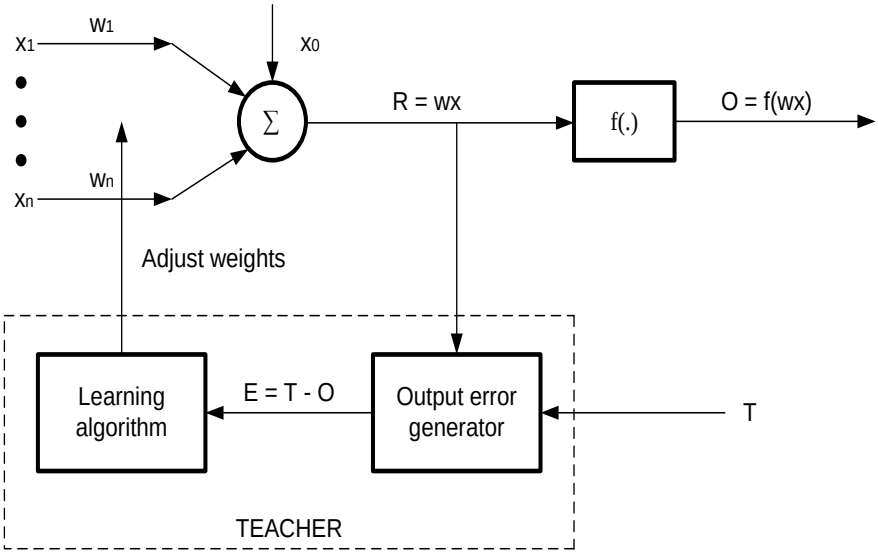


Fig 1. Model of a neuron with error-correcting learning

We start with the least-squared error (E) between the target output (T) and actual output (O) defined by [7]:

$$E = \frac{1}{2}(T_i - O_i)^2 = \frac{1}{2}[T_i - f(w_i, x_i)]^2, \qquad (5)$$

where $\boldsymbol{T}$ – *target output,*

$\boldsymbol{O}$ – *actual output,*

w_i – *weight,*

x_i, O_i, T_i – *the vector representations of the input and target output, respectively of the i neuron,*

f – *activation function.*

In proposed model activation function is nonlinearity defined by:

$$f(R) = (1 + e^{-kR})^{-1}, \quad (6)$$

$$k\epsilon[-\infty, +\infty],$$

$$R = w_i(k+1)x_i \quad (7)$$

The error gradient vector is:

$$\nabla E = -(T_i - O_i)\, f'(w_i, x_i)x_i \quad (8)$$

Since we are seeking minimization of the error we consider:

$$\Delta w_i = -\mu \nabla E,$$

where μ is a positive constant.

Hence Δw_i become:

$$\Delta w_i = \mu(T_i - O_i)\, f'(w_i, x_i)x_i \quad (9)$$

Applying discrete mathematics, we update the weight vector, using:

$$w_i(k+1) = w_i(k) + \mu(T_i - O_i)\, f'(w_i, x_i)x_i \quad (10)$$

We assume that the sigmoid nonlinearity with a constant k= 1 is used. The gradient is approximated by the relation:

$$f'(w_i, x_i) = \frac{1}{2}(T_i - O_i^2) \quad (11)$$

$$w_i(k+1) = w_i(k) + \frac{\mu}{2}(T_i - O_i)\left(T_i - O_i^{\,2}\right)x_i \quad (12)$$

With the new vector we calculate the output signal and if the error is the larger than a desired value, repeat the above problem until the desired minimum error is reached.

Model

Figure 2 represents here proposed hybrid model, combining two methods.

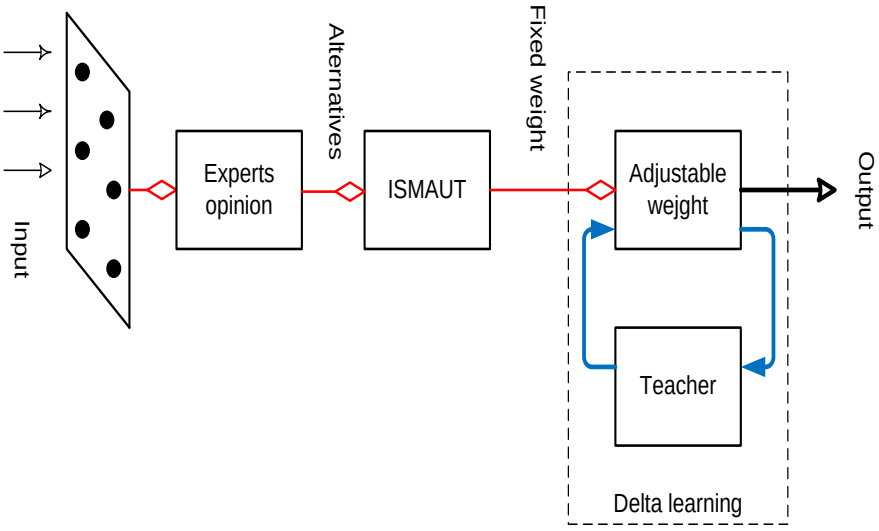


Fig 2. Hybrid model

Application

The following example illustrates how implicit preference relations can be derived.

Table 1: Feasible alternatives for assessing of risk events and incidents.

Alternative	*v1(r1)*	*v2 (r2)*	*v3(r3)*	*v4(r4)*	*v5(r5)*	*v6(r6)*	*v7(r7)*	*v8(r8)*
a1	0.5	0.3	0.8	0.1	0.8	0.1	0.0	0.2
a2	0.4	0.1	0.7	0.3	0.6	0.7	0.2	0.3
a3	0.3	0.2	0.6	0.2	0.7	0.4	0.1	0.2
a4	0.6	0.4	0.9	0.3	0.5	0.8	0.5	0.1
a5	0.3	0.0	0.8	0.5	0.4	0.2	0.3	0.6
a6	0.8	0.3	0.5	0.2	0.5	0.4	0.5	0.2
a7	0.2	0.5	0.6	0.4	0.9	0.7	0.3	0.5

Information security events and incidents are:

- ***v1*** - loss of service, equipment or facilities,
- ***v2*** - system malfunctions or overloads,
- ***v3*** - human errors,
- ***v4*** - non-compliances with policies or guidelines,
- ***v5*** - breaches of physical security arrangements,
- ***v6*** - uncontrolled system changes,
- ***v7*** - malfunctions of software or hardware,
- ***v8*** - access violations.

$$a_k = \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$$

The alternatives in Table 1 present the opinion of experts. After applying the method ISMAUT the weights are:

$$w = \begin{bmatrix} 0.1995902 \\ -0.5372951 \\ 0.0147541 \\ 1 \\ 0.8459016 \\ -0.1913934 \\ 0.3114754 \\ -0.6430328 \end{bmatrix}$$

Results

First compare any two alternatives derived from expert opinions. In our case it is difficult to identify preferred alternatives and we compare any two alternatives. For the alternatives described in Table 1, we have seven comparisons. Moreover, the sum of the weights must be one and this is the eighth equation.

To determine the weight vectors we solve the following system of equations:

1. $[v_1(r_1) - v_1(r_2)]w_1 + [v_2(r_1) - v_2(r_2)]w_2 + [v_3(r_1) - v_3(r_2)]w_3 + [v_4(r_1) - v_4(r_2)]w_4 + [v_5(r_1) - v_5(r_2)]w_5 + [v_6(r_1) - v_6(r_2)]w_6 + [v_7(r_1) - v_7(r_2)]w_7 + [v_8(r_1) - v_8(r_2)]w_8 = 0$

2. $[v_1(r_2) - v_1(r_3)]w_1 + [v_2(r_2) - v_2(r_3)]w_2 + [v_3(r_2) - v_3(r_3)]w_3 + [v_4(r_2) - v_4(r_3)]w_4 + [v_5(r_2) - v_5(r_3)]w_5 + [v_6(r_2) - v_6(r_3)]w_6 + [v_7(r_2) - v_7(r_3)]w_7 + [v_8(r_2) - v_8(r_3)]w_8 = 0$

3. $[v_1(r_3) - v_1(r_4)]w_1 + [v_2(r_3) - v_2(r_4)]w_2 + [v_3(r_3) - v_3(r_4)]w_3 + [v_4(r_3) - v_4(r_4)]w_4 + [v_5(r_3) - v_5(r_4)]w_5 + [v_6(r_3) - v_6(r_4)]w_6 + [v_7(r_3) - v_7(r_4)]w_7 + [v_8(r_3) - v_8(r_4)]w_8 = 0$

4. $[v_1(r_4) - v_1(r_5)]w_1 + [v_2(r_4) - v_2(r_5)]w_2 + [v_3(r_4) - v_3(r_5)]w_3 + [v_4(r_4) - v_4(r_5)]w_4 + [v_5(r_4) - v_5(r_5)]w_5 + [v_6(r_4) - v_6(r_5)]w_6 + [v_7(r_4) - v_7(r_5)]w_7 + [v_8(r_4) - v_8(r_5)]w_8 = 0$

5. $[v_1(r_5) - v_1(r_6)]w_1 + [v_2(r_5) - v_2(r_6)]w_2 + [v_3(r_5) - v_3(r_6)]w_3 + [v_4(r_5) - v_4(r_6)]w_4 + [v_5(r_5) - v_5(r_6)]w_5 + [v_6(r_5) - v_6(r_6)]w_6 + [v_7(r_5) - v_7(r_6)]w_7 + [v_8(r_5) - v_8(r_6)]w_8 = 0$

6. $[v_1(r_6) - v_1(r_7)]w_1 + [v_2(r_6) - v_2(r_7)]w_2 + [v_3(r_6) - v_3(r_7)]w_3 + [v_4(r_6) - v_4(r_7)]w_4 + [v_5(r_6) - v_5(r_7)]w_5 + [v_6(r_6) - v_6(r_7)]w_6 + [v_7(r_6) - v_7(r_7)]w_7 + [v_8(r_6) - v_8(r_7)]w_8 = 0$

7. $[v_1(r_7) - v_1(r_1)]w_1 + [v_2(r_7) - v_2(r_1)]w_2 + [v_3(r_7) - v_3(r_1)]w_3 + [v_4(r_7) - v_4(r_1)]w_4 + [v_5(r_7) - v_5(r_1)]w_5 + [v_6(r_7) - v_6(r_1)]w_6 + [v_7(r_7) - v_7(r_1)]w_7 + [v_8(r_7) - v_8(r_1)]w_8 = 0$

8. $w_1 + w_2 + w_3 + w_4 + w_5 + w_6 + w_7 + w_8 = 1$

We replace $v_i(r_j)$ with results expert assessment of risk objects (*Table 1.)*:

1. $[0.5 - 0.4]w_1 + [0.3 - 0.1]w_2 + [0.8 - 0.7]w_3 + [0.1 - 0.3]w_4 + [0.8 - 0.6]w_5 + [0.1 - 0.7]w_6 + [0 - 0.2]w_7 + [0.2 - 0.3]w_8 =$

2. $[0.4 - 0.3]w_1 + [0.1 - 0.2]w_2 + [0.7 - 0.6]w_3 + [0.3 - 0.2]w_4 + [0.6 - 0.7]w_5 + [0.7 - 0.4]w_6 + [0.2 - 0.1]w_7 + [0.3 - 0.2]w_8 = 0$

3. $[0.3 - 0.6]w_1 + [0.2 - 0.4]w_2 + [0.6 - 0.9]w_3 + [0.2 - 0.3]w_4 + [0.7 - 0.5]w_5 + [0.4 - 0.8]w_6 + [0.1 - 0.5]w_7 + [0.2 - 0.1]w_8 = 0$

4. $[0.6 - 0.3]w_1 + [0.4 - 0]w_2 + [0.9 - 0.8]w_3 + [0.3 - 0.5]w_4 + [0.5 - 0.4]w_5 + [0.8 - 0.2]w_6 + [0.5. - 0.3]w_7 + [0.1 - 0.6]w_8 = 0$

5. $[0.3 - 0.8]w_1 + [0 - 0.3]w_2 + [0.8 - 0.5]w_3 + [0.5 - 0.2]w_4 + [0.4 - 0.5]w_5 + [0.2 - 0.4]w_6 + [0.3 - 0.5]w_7 + [0.6 - 0.2]w_8 = 0$

6. $[0.8 - 0.2]w_1 + [0.3 - 0.5]w_2 + [0.5 - 0.6]w_3 + [0.2 - 0.4]w_4 + [0.5 - 0.9]w_5 + [0.4 - 0.7]w_6 + [0.5 - 0.3]w_7 + [0.2 - 0.5]w_8 = 0$

7. $[0.2-0.5]w_1 + [0.5-0.3]w_2 + [0.6-0.8]w_3 + [0.4-0.1]w_4 + [0.9-0.8]w_5 + [0.7-0.1]w_6 + [0.3-0]w_7 + [0.5-0.2]w_8 = 0$

8. $w_1 + w_2 + w_3 + w_4 + w_5 + w_6 + w_7 + w_8 = 1$

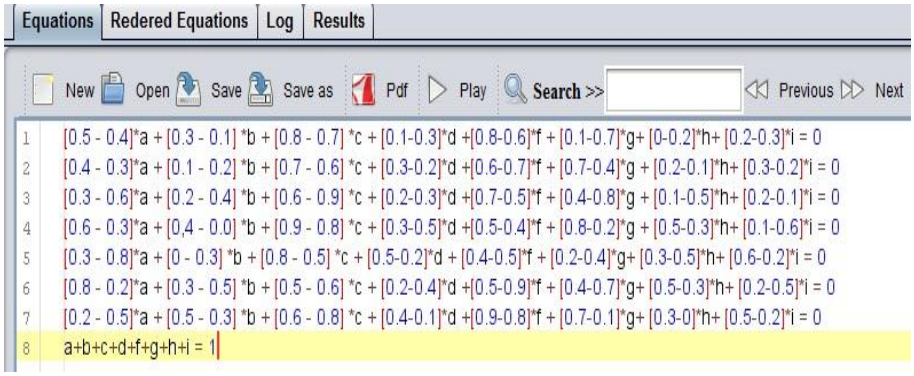


Fig. 3. Calculation of weights

After calculations we receive the weights:

$$w_1 = 0.1995902,\ w_2 = -0.5372951,\ w_3 = 0.0147541,\ w_4 = 1,$$
$$w_5 = 0.8459016,\ w_6 = -0.1913934,\ w_7 = 0.3114754,\ w_8 = -0.6430328.$$

Fig. 4. Result of calculation of weights

Before training the Perceptron we check the following:

- Choose a set of input vectors {x} – first alternative.
- Choose a set of target output {T} – 0.0.
- Choose a small positive value of the learning rate. μ = 1.
- Select the gain (a constant of the function). k = 1
- Define a small acceptable error value for the actual output. When this error is reached, the iteration process terminated. Another termination procedure sets the maximum number of iterations.

$$w(1)x = [0.5\ \ 0.3\ \ 0.8\ \ 0.1\ \ 0.8\ \ 0.1\ \ 0.0\ \ 0.2]\begin{bmatrix} 0.1995902 \\ -0.5372951 \\ 0.0147541 \\ 1 \\ 0.8459016 \\ -0.1913934 \\ 0.3114754 \\ -0.6430328 \end{bmatrix} = 0.579$$

$$O(1) = f(0.579) = \frac{1}{[1 + 1/exp(0.579)]} = 0.641$$

$$f'(w(1)x_i) = \frac{1}{2}\left[T_i - \left(O_i(1)\right)^2\right] = \frac{1}{2}[0 - (0.641)^2] = -0.205$$

$$w(2) = w(1) + \mu\left(T - O(1)\right) f'(w(1)x_i)x_i$$

Result of this iteration is given in figure 5.

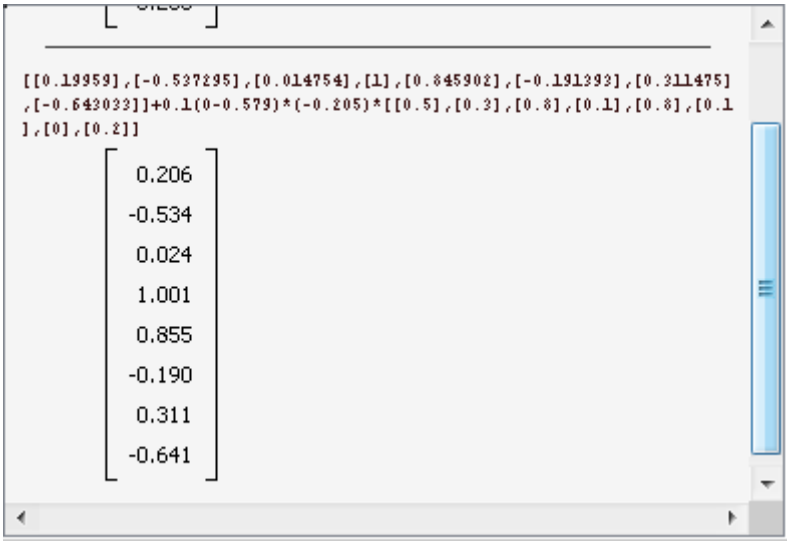


Fig. 5. Adjust weights

$$E = \frac{1}{2}(T_i - O_i)^2 = \frac{1}{2}[T_i - f(w_i, x_i)]^2 = \frac{1}{2}[0 - f(0.579)]^2 = \frac{1}{2}[-0.641]^2 = 0.205$$

In our example we obtain an acceptable error and the procedure to minimize the error ends.

Summary

The evaluation is based on Imprecisely Specified Multi-Attribute Utility Theory, which in contrast to standard utility theory, does not require complete enumeration of all specified alternatives. This is important since most decision making problems are too complex to allow complete enumeration of all feasible alternatives. With Delta learning algorithm we get adjusts weighs and this way minimize the risk events impact. The proposed approach is applicable in many different cases of risk assessment and allows the evaluation of quantitative measure of anticipated risk.This instruction file for Word users (there is a separate instruction file for LaTeX users) may be used as a template. Kindly send the final and checked Word and PDF files of your paper to the Contact Volume Editor. This is usually one of the organizers of the conference. You should make sure that the Word and the PDF files are identical and correct and that only one version of your paper is sent. It is not possible to update files at a later stage. Please note that we do not need the printed paper.

We would like to draw your attention to the fact that it is not possible to modify a paper in any way, once it has been published. This applies to both the printed book and the online version of the publication. Every detail, including the order of the names of the authors, should be checked before the paper is sent to the Volume Editors.

References

1. Andriole Stephen, Advanced Technology for Command and Control Systems Engineering, November 1990.
2. Gardner, M.W., Dorling S.R., Artificial Neural Networks (the multilayer perceptron) – a review of applications in the atmospheric sciences, Atmosphere Environment, 1999, pp. 709-719.
3. P-ISO 27001-2005.
4. ISO 17799-2005.
5. D'Ambrosio, Joseph G., ISMAUT Tools: A Software Tool Kit for Rational Tradeoffs Among Conflicting Objectives 1996.
6. Holzmann, Gerard J., Early Fault Detection Tools, Tools and Algorithms for the Construction and Analysis of Systems, 2nd International Workshop, TACAS'96, Passau, Germany, 1996, pp. 1-13.
7. Take Fuji Y., Holis P., Foo Y.P., Cho Y.B., Error Correcting System Based on Neural Circuits, 1st International Conference on Neural Networks, vol. 3, 1987, pp. 293-300.
8. Kartalopoulos Stamatios, Understanding Neural Networks and Fuzzy Logic, 1996.
9. DoDAF_v1_Deskbook, 2004.

Exposing the Intellectual Assets of a University Department

Stanislav Ivanov[1], Juliana Peneva[1]

[1] New Bulgarian University, Department of Computer Science, Montevideo str. 21, 1618 Sofia, Bulgaria
{Juliana Peneva, jpeneva@nbu.bg ,
Stanislav Ivanov, sivanov@nbu.bg

Abstract. Works for designing and deploying a university department repository are presented. General considerations about policies and functional requirements are outlined with respect to the institution context. Technical and system issues are briefly discussed. Finally the benefits and usability of the repository are summarized.

Keywords: H3.5 On-line information services– Data sharing; H3.7 Digital libraries-Collection

1 Introduction

The proper management of resources is fundamental for any organization in response to changing business conditions. Many companies are achieving a business advantage by managing successfully their digital resources. Resources include documents, images, video or audio clips, animations, presentations, online courses, web pages, etc. Organizations are of different types and sizes ranging from SME to international corporations. All of them demonstrate an intensive use of digital resources, the resources being stored, distributed, shared and reused without difficulty.

Nowadays digital resources are increasingly being recognized as a very important organizational asset au par with finance and human resources. The business managerial and technical benefits of digital resources are summarized in [1]. In order to examine their value [2] and to consider the opportunities for reuse digital resources are organized in repositories that support the organizations' policy on digital asset management.

For higher education institutions digital resources can be used strategically to expose intellectual assets such as learning content and research papers. Because of the very rapid development in information and communications technologies it becomes possible to manage the university assets more effectively than it the past two decades. Digital resources are organized as university-based institutional repositories [3], including long-term preservation and distribution.

During the last five years different types of repositories ranging from digital libraries through various institutional collections and e-journals up to collaborative learning environments have been built. In this perspective, universities and scientific institutions demonstrate a remarkable activity. The main reason for this perpetual activity is the huge diversity of purposes, deposited resources, services and potential users. Universities need to exhibit and deploy different kinds of their intellectual assets. It is a matter not only of user's convenience, but of representativeness and prestige as well. In this plan, it is quite natural

that the main share of active repositories belongs to countries with advanced higher education and science.

In the dynamic domain of computer science it appears especially helpful to store many useful products of the educational process itself: the so-called grey literature e.g. case studies, student's research projects, diploma theses, working papers, technical reports, etc. Active learning could be significantly stimulated via a proper repository as well. This determines our decision to develop our own, at department level, digital repository to deploy digital content not covered by the university information structures: LMS Moodle and Scholar electronic repository. Our development should deliver digital materials not offered by these two systems. Ensuring that proper digital materials are long term visible is very important for the department as part of its positioning strategy. The goal of this repository would be to provide added value to the Computer Science Education community, to our students and alumni. Moreover, the university educational policy encourages the shift towards e-learning and a flexible learning process. This implies reducing the face-to-face sessions, disseminating online coursework on a wider basis and training the students any time. So, designing a new infrastructure project and applying a standards-based approach to the management, preservation and access of existing and future digital resources is essential for the department to fulfill its mission as a team of lecturers and researchers.

In the context of the above, the main goal of this paper is to present our work on designing an institutional repository of the Department of Informatics at New Bulgarian University. We discuss what do we need and determine the type of the material to be stored in the repository. Creating a proper digital collection that captures and preserves the department's intellectual output would increase its visibility and public value. This repository will support learning and administrative processes of our department. To build an effective repository the technical set up process is to be planned properly. The paper presents general considerations about policies and functional requirements with respect to the institution context. Technical and system issues are also briefly discussed. Finally the benefits of the repository are summarized and its usability is justified.

2 Institutional repositories in Bulgaria

The Bulgarian research community is aware about the benefits of exposing research content in digital form. Therefore in recent years open access repositories ranging from digital libraries up to e-journals have been built. Large companies are reporting for own repository investigations as well. In addition there are many workshops and annual open repositories [4] conferences that concentrate on important issues concerning repository creation and management. Despite of the disappointments for many organizations due to the resulted greater than expected costs for set up a repository, research effort in this area appears promising. Repositories increase successfully very quickly. In this perspective, universities and scientific institutions demonstrate a remarkable activity. Open access academic repositories marked a boost of 300 during the mid of 2006. Since the beginning of year 2007 the growth of such repositories listed in the *Open*DOAR Database [5] shows a constant increase of 300 repositories per year up to its present number of about 2200 (Fig.1).

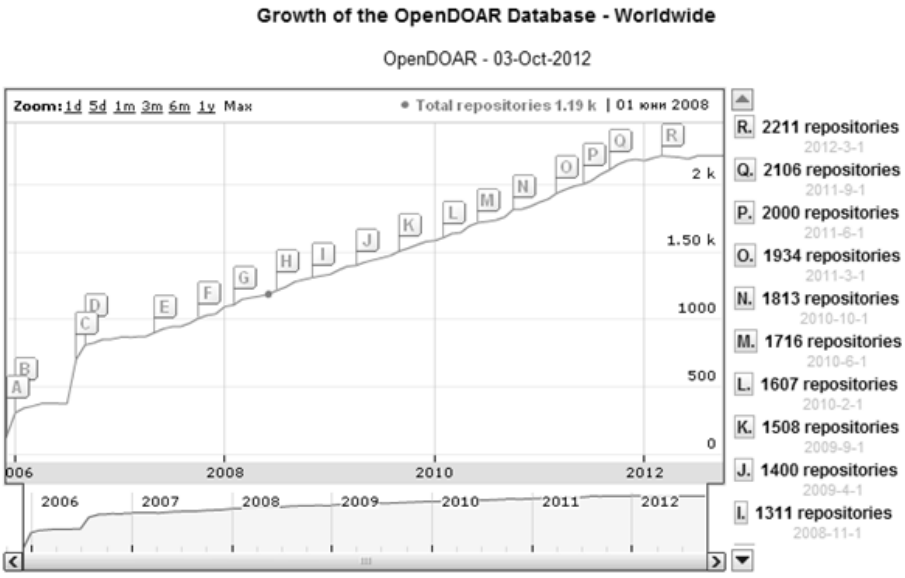


Fig. 1 Growth of the OpenDOAR Database – Worldwide - 04-Nov-2012

Up to now about 2200 scholar repositories all over the world have been reported, about 20% of them in the Nord America, 47% in Europe shared among United Kingdom, Germany, Spain, France and Italy. Other 18% reside in Asia, about 6% - in South America. In Bulgaria there are six open repositories (Fig.2) only [6, 7, 8, 9, 10, 11] in the *Open*DOAR Database [5].

3 The institutional repository of the Department of Informatics

According to the SPARC alliance [12] institutionally defined repositories are scholarly, cumulative, open and interoperable. Generally speaking a departmental repository can be compared to a database with a set of services used to store, index and preserve scholarly materials, research findings etc. in digital formats. The main goal is to manage and disseminate digital materials created by the department and its community members [13].

Organization	**Description**	**Software**	**Subjects**	**Content/ Items**
Academic research repository at Burgas Free University	research output	DSpace	Computers and IT; Arts and Humanities General; Business and Economics; Law and Politics	Articles; Conferences; Books/ 226
DSpace at IMI	journal output	DSpace	Multidisciplinary	Articles; Conferences; Books/ 1652

Bulgarian OpenAIRE Repository	peer-reviewed publications FP7 projects	DSpace	Multidisciplinary	Books/ 107
Electronic Repository-Central Medical Library-Sofia	Research output	DSpace	Multidisciplinary; Health and Medicine	Articles; Theses; Books; Learning Objects/ 155
New Bulgarian University Scholar Electronic Repository	Research output	EPrints	Multidisciplinary	Articles; Conferences; Books; Learning Objects/ 899
Research at Sofia University	Research output	DSpace	Multidisciplinary	Articles/ 555

Fig. 2 Open Repositories in Bulgaria - 04-Nov-2012

The repository will be used for electronic publishing and housing of different digitized collections concerning the knowledge management of the department. The final goal is to offer open access to scholarly research. So, the departmental repository is designed to deploy content not managed via university learning facilities: investigations, learning resources, theses, students' projects and papers. The goals are twofold:

- to store and represent the department's intellectual production;
- to prevent the student's plagiarism making visible the majority of the coursework done during the years.

We follow the guidelines given in [14] for each stage of building the repository bearing in mind the requirements of our institution about copyrights, access rights etc.

Steps in building the repository:

1. Determination of the content type.

We will accept bachelor, masters and doctoral theses, student's research materials and original learning content from the department of Informatics. The user will not be allowed to download copyright protected content.

2. Service definition

Repository services concerns the management of corpora i.e. annotated collections of digitized objects. Making visible the stored content to the user groups can be defined as a top service priority. The service's mission is to raise the visibility of the Department of Informatics at New Bulgarian University. It will facilitate our students, extending their access to properly collected and organized additional learning materials. Key users of the departmental repository are going to be students and faculty.

3. Policy framework

A policy framework is very important to determine the operational boundaries within which the repository will deliver its services. This framework contributes for an easier use of the repository, permits for it support and facilitates the decision-making processes. Some

policies need legal agreements i.e. definition of a deposit license and usage license that user agree to.

Policies can be classified as strategic and operational. Strategic policies reflect the wider strategic policies of the institution. New Bulgarian University has a high- profile vision statements [15] and defined procedures concerning research, teaching and theses. Following them the repository can be easily embedded within the university. Administrators will survey the deposit of diploma theses and other research output. As learning and teaching materials are deposited within Moodle, their store in the repository is optional.

Operational policies deal with day-to-day operations. They comprise:

- Submission policies – only administrators will be allowed to deposit submitted materials after approval.
- Collection policies – the repository will focus on computer science and mathematics. Final reviewed/assessed versions will be accepted only.
- Preservation policies – different policies will be set for different type of materials. We will keep theses as deposited whilst teaching materials, because of the dynamics in the computer science area more likely will be updated. Regular backups will be made, the frequency being adjusted upon the intensity of use.

4. Choice of an infrastructure

Taking into account that flexibility among the different collections is a key feature the goal of the department repository is to offer a proper infrastructure with a well defined range of services. A high level archival model to act as a framework is necessary. We adopt a well established model in this area – OAIS (Reference Model for an Open Archival Information System) [16] –see Fig.3.

The OAIS environment is made up of the OAIS, i.e. the digital library system, the producers and consumers of its content and services, and the management and strategic input into the system. Within the OAIS are six main functions: ingest (submit), data management, storage, access, administration and planning. Common services e.g. operating system and networking services are assumed to be available. Evaluations concerning the usability of OAIS to build different kind of digital repositories are given in [17].

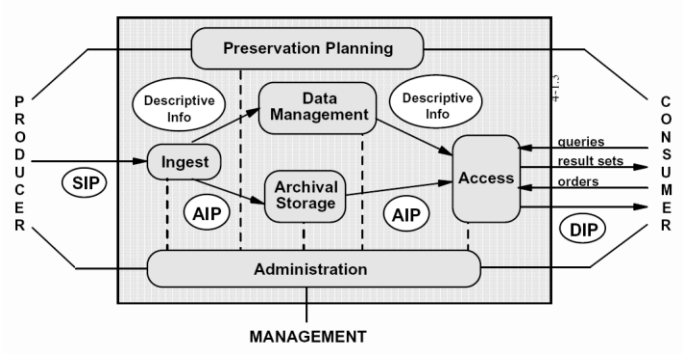


Fig.3 OAIS Functional Entities (from Reference Model for an Open Archival Information System - OAIS, 2009, Fig 4-1)

5. Investigation of possible repository solutions

Taking into account the requirements we can make decisions concerning the repository infrastructure. To set up a repository three approaches can be followed: do-it-yourself; use standard packages and outsourcing (external hosting). With limited staff resources for long-term maintenance and support we have chosen to apply the most popular approach i.e. to use a standard package nevertheless that external hosting becomes recently more popular. A wide variety of available software with different features and strengths exists. Recently the more commonly adopted software solutions fall into two broad groups: open source and commercial software. Our investigations show that there are over 316 repositories using the EPrints software, about 901 – DSpace,95 – Digital Commons. The rest of the software exhibits a limited application. EPrints is an open source platform for building repositories of documents like research literature, scientific data, and student theses. DSpace preserves and enables easy and open access to all types of digital content including text, images, moving images, mpegs and data sets. It is applied for accessing, managing and preserving scholarly works. Digital Commons offers external hosting for institutional repositories. It can include pre-prints and/or final copies of working papers, journal articles, dissertations, master's theses, conference proceedings, and a wide variety of other content types.

6. Pilot testing

The next stage of the process of implementing a digital repository for the needs of our department was to select a system, which could be the most suitable for our needs. So, following the local policies and practices we have decided to run an open source repository platform. This choice reflects the good will and the IT expertise of the department's staff. Running open source software appears to be the cheapest solution as the installation and the customization of the repository require a relatively short list of intensive activities. The skills required depend on specific repository platforms i.e. the programming language they are written in. There are common skills such as HTML, Web page design, SQL applicable to all choices. In order to justify the choice of the software, pilot installations of some open source packages has been undertaken. These were used as test beds for the overall repository development. A pilot system was used to tune the software parameters.

We made a short-list, where the exclusionary criteria were popularity and price. Even with a short-list consisting of only three systems, Dspace, Eprints and Digital Commons, the list of criteria we considered relevant for evaluating these systems was too long. Thus in order to make a more quantifiable and formal judgment we decided to formulate the problem as a multiple criteria selection problem and to use a decision support system to solve it. The details of the overall process are presented in [18] and lie beyond the scope of this paper. The departmental repository has been implemented using EPrints because the system is customizable and branding-friendly. Respecting the university policy about the diploma theses visibility the repository is accessible at http://infdep.nbu.bg, but only at local level from the university buildings.

4 Conclusion

The department repository provides thematically oriented texts – results of our students' work during various educational activities. The decision to create one more repository to manage proper digital content is challenging. One could argue that organizational digital assets already are stored in many types of systems e.g. locally

developed closed systems, virtual learning environments, portals, etc. However the storage of students' scholar work facilitates the control on plagiarism. It is up to the department to track a direct borrowing of texts and in a broader sense the plagiarism of themes, ideas, methods and approaches. In addition active learning is stimulated.

Summarizing the benefit is:

- overcoming temptation for cheating
- availability of quality examples of students' achievements
- a comparative analysis of students' work can be performed

As a result, we expect a significant increase of the quality of students' paper work: theses, research projects, surveys, etc.

References

1. Duncan C. Digital Object Repositories Explained, an Intrallect White Paper, 2006.
2. Bluthe E., Chandra V. The Value Proposition in Institutional Repositories EDUCASE Review, September/October 2005.
3. Crow R. The Case for Institutional Repositories: A SPARC Position Pape SPARC, 2002.
4. http://openrepositories.org/
5. www.opendoar.org
6. http://eprints.nbu.bg/ – New Bulgarian University Scholar Electronic Repository
7. http://research.it.fmi.uni-sofia.bg - Research at Sofia University
8. http://sci-gems.math.bas.bg - DSpace at IMI
9. http://nt-cmb.medun.acad.bg:8080/jspui/ - Central Medical Library, Medical University, Sofia
10. http://www.bg-openaire.eu/ - Bulgarian OpenAIRE Repository.
11. http://research.bfu.bg:8080/jspui/ - Academic Research Repository at the Burgas Free University.
12. http://www.arl.org/sparc
13. Lynch C. "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age" ARL, no. 226 (February 2003): 1-7.
14. http://ciard.net/pathways - Develop a Repository for Digital Content, October, 2009.
15. http://www.nbu.bg
16. CCSDS 650.0-P-1.1, Reference Model for an Open Archival Information System (OAIS) (Pink Book, Issue 1.1, August 2009)
17. Allinson J et al. OAIS as a reference model for repositories, UKOLN, University of Bath, 2006.
18. Andonov F. et al. A Digital Repository at University Department Level The 7 Annual International Conference on Computer Science and Education in Computer Science, July 06-10 2011, Sofia, Bulgaria.

SYSTEM FOR MONITORING OF THE UNIVERSITY COMPUTER NETWORK PERFORMANCE

Dimitar Dimitrov
VTU - Sofia

Abstract. This article presents the results of development of an information system to monitor and control the operation of the university system for distribution of Internet services. The approach used is based on the application of a standard deployment tools to monitor and control the operation of devices and information services

Keywords: Internet, systems, services, monitoring, cable and wireless computer network, campus.

1 Introduction

Modern information systems require providing even greater capacity both of the main Internet highways as well as the broadband access for the end users. Nowadays it is not enough just to digitalize and transfer units of information, but is usually in the form of audio and video transmissions in so-called online-meeting system, that is actually "live" meetings in the network.

This includes so-called information flows of different kinds (streaming TV programs, money transfers and many more). In universities and other academic and educational institutions there are specific information flows dedicated to educational needs.

This paper presents the main features of the renovated and updated Local information network and Internet access on the territory of the University of Transport campus in Sofia. This information network consists of two main component: optic wire system and wireless connectivity characterized by high capacity data transmission, and wireless mobility.

The main focus of this paper is the embedded program for complex monitoring of system performance, and the technical means and methods for monitoring and reporting of network problems. This program is an adapted and configured version the popular open-source system NAGIOS, and has been successfully implemented with very good performance for the last four months

The establishment of this monitoring system lead to improved service quality and successful trouble shooting.

2 TECHNICAL PARAMETERS OF THE SYSTEM

Computer networking is a two-component modular nature and contains an internal wired and wireless access terminal. Two optical communication channels provide comprehensive IT services, including the delivery of telephone and Internet [1,2].The main components of the renovated and built a network are:

-optic channel and the internal network connectivity to the Internet with a capacity of 250 Mbps;
-optic network connectivity between buildings with a capacity of 1 Gbps;
-internal copper-network connectivity with capacity of 1 Gbps, in the endpoints is provided capacity of 100 Mbps;
-enclosed and open area wireless connectivity with WiFi type 100% secured capacity: 802.11n Draft: up to 300 Mbps; 802.11g: 6, 9, 12, 18, 24, 36, 48, 54Mbps; 802.11b: 1, 2, 5.5, 11Mbps;
-smart network equipment 10/100/1000 Mbps (SMC6128L2);
-WiFi network equipment and antennas (Wireless N Gbit Router ASUS RT-N16, 300Mbps, 2xUSB, Linksys WRT-54GL, EnGenius EAP-9550);
-supporting routers to connect to the internet (CISCO and MikroTik).
-transition module Audio Codes Mediant-600 to the corporate PBX Alcatel 4400 to provide telephone communications over the Internet.

In Fig. 1 presents the general model of the information network. The network itself incorporates multiple coverage with over 1 200 points of support for wired access, and it is possible to further increase their number. Wireless WiFi connectivity is achieved by installation of wireless devices, optimal distribution scheme, thereby ensuring a high level of coverage and mobility of the signal.

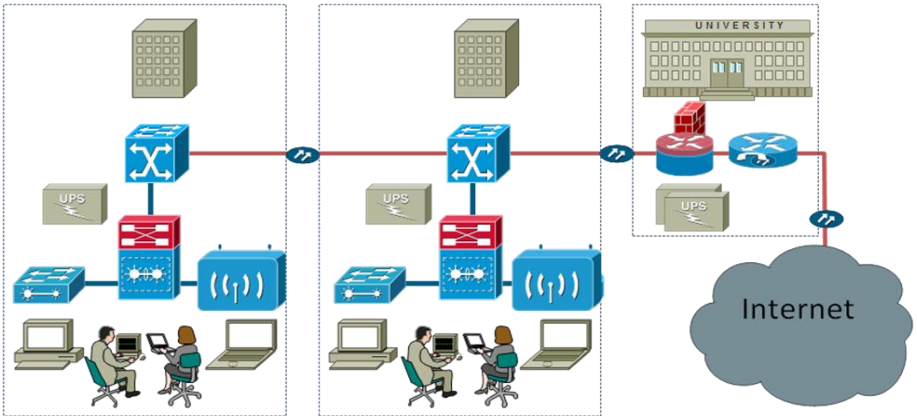


Fig. 1. View the general network system model

The system consists of the required number of devices and allows for multiple software setup and calibration characteristics for effective traffic management and control of data flow.

3 SYSTEM PERFORMANCE AND MONITORING

Logical connectivity of the network corresponds to the physical infrastructure of the network. Access control and logical deployment is based on the three lowest layers of access (the famous OSI model):
-monitoring and control of the availability of physical access (Layer 1);

-access control of the data exchange level (Layer 2) by means of recording the physical (MAC) addresses and the use of protection in terms of duplication of address spaces, DHCP spoofing server, elaboration and implementation of and use of logical network separation model VLANs, etc.
-access control on the network level (Layer 3), providing the TCP / IP v4 model of dynamic virtual network.

The use of such structural separation and logical segmentation allows for reduction of collisions in the network caused by faulty settings and configurations of end users (computers). We Used high-end devices, which allow for multiple software setup and calibration characteristics for effective traffic management and control of data flow.

Use the supplied tools Nagios [3], which is a powerful monitoring system allows you to identify and resolve infrastructure problems. The system after discontinuation of the network and take measures to correct the problems of users. This system provides the following options:
-plan for infrastructure upgrades before outdated systems cause failures;
-respond to issues at the first sign of a problem;
-automatically fix problems when they are detected;
-coordinate technical team responses;
-ensure your organization's SLAs are being met;
-ensure IT infrastructure outages have a minimal effect on your organization's bottom line;
-monitor your entire infrastructure and business processes.

In the main communications center we have a dedicated server that has installed on it the specialized software Nagios. The system is set up to monitor the devices in the following groups:
-network routers;
-network switches;
-Wi-Fi devices;
-Linux servers
-Windows servers and critical machines;
-video IP or DVR cameras and etc.

The system has been configured to monitor critical IT infrastructure components, including system metrics, network protocols, applications, services, servers, and network infrastructure. The system includes some network protocols, applications and services, such as:
-uptime;
-ping check by analyzing packet loosing;
-checking memory usage;
-checking usage storage space;
-checking CPU performance;
-checking common services ssh, www, mail and etc.

Nagios sends alerts when critical infrastructure components fail and recover, providing administrators with notice of important events. Alerts can be delivered via email, SMS, or custom script.

Reports provide a historical record of outages, events, notifications, and alert response for later review. Availability reports help ensure your SLAs are being met.

Figures 2, 3, 4 present the main screen of the Nagios monitor system.

The displayed screens give a basic idea of the capabilities of the monitoring system for the inspection and control access to all network devices, servers and services. Performance of this system is connected with the online registration system log and alert problems.

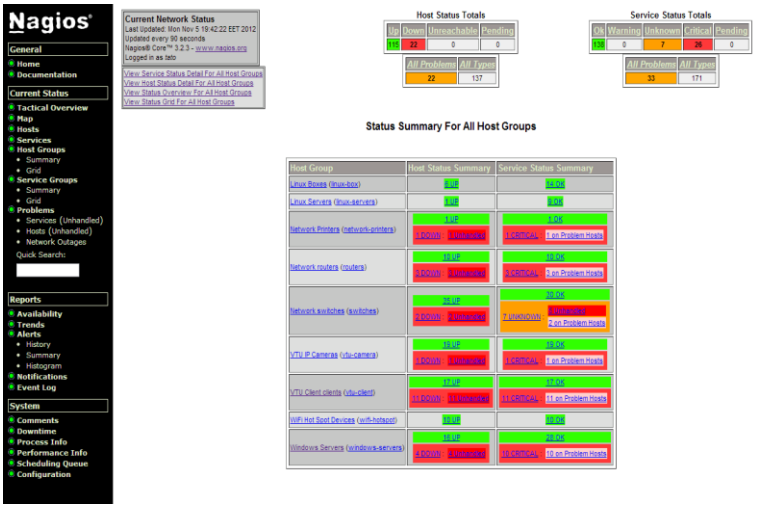


Fig. 2. View Status Overview for All Host Groups

Host	Service	Status	Last Check	Duration	Attempt	Status Information
a4400-pc	PING	OK	11-05-2012 20:22:32	25d 6h 10m 24s	1/4	PING OK - Packet loss = 0%, RTA = 0.30 ms
ad1.vtu.local	PING	OK	11-05-2012 20:23:49	25d 18h 36m 46s	1/4	PING OK - Packet loss = 0%, RTA = 0.36 ms
ad2.vtu.local	C:\ Drive Space	CRITICAL	11-05-2012 20:23:44	109d 22h 37m 9s	3/3	No route to host
	CPU Load	CRITICAL	11-05-2012 20:23:04	109d 22h 43m 22s	3/3	No route to host
	Memory Usage	CRITICAL	11-05-2012 20:17:16	109d 22h 37m 4s	3/3	No route to host
	NSClient++ Version	CRITICAL	11-05-2012 20:23:04	109d 22h 43m 18s	3/3	No route to host
	PING	CRITICAL	11-05-2012 20:19:25	109d 22h 37m 0s	3/3	CRITICAL - Host Unreachable (172.16.5.5)
	Uptime	CRITICAL	11-05-2012 20:23:04	109d 22h 43m 13s	3/3	No route to host
	W3SVC	CRITICAL	11-05-2012 20:19:25	109d 22h 36m 55s	3/3	No route to host
alcatel-pc	PING	OK	11-05-2012 20:20:46	66d 2h 59m 30s	1/4	PING OK - Packet loss = 0%, RTA = 0.21 ms
alfa-group-sc2	PING	OK	11-05-2012 20:22:32	45d 23h 38m 52s	1/3	PING OK - Packet loss = 0%, RTA = 1.12 ms
ap-chkolev	PING	OK	11-05-2012 20:19:57	0d 9h 25m 17s	1/3	PING OK - Packet loss = 0%, RTA = 3.28 ms
ap-rt-n16-k1	PING	OK	11-05-2012 20:20:42	0d 9h 34m 20s	1/3	PING OK - Packet loss = 0%, RTA = 0.40 ms
ap-rt-n16-k3	PING	OK	11-05-2012 20:22:16	0d 9h 38m 16s	1/3	PING OK - Packet loss = 0%, RTA = 0.41 ms
ap-rt-n16-k4	PING	OK	11-05-2012 20:22:32	23d 0h 20m 5s	1/3	PING OK - Packet loss = 0%, RTA = 3.82 ms
ap-rt-n16-so1-bb	PING	OK	11-05-2012 20:23:57	0d 0h 56m 5s	1/3	PING OK - Packet loss = 0%, RTA = 0.87 ms
ap-rt-n16-so1-stol	PING	OK	11-05-2012 20:22:57	0d 0h 57m 5s	1/3	PING OK - Packet loss = 0%, RTA = 2.34 ms
ap-rt-n16-so2-pilon	PING	OK	11-05-2012 20:20:46	42d 1h 45m 35s	1/3	PING OK - Packet loss = 0%, RTA = 3.33 ms
ap-wl-520gu-r24	PING	OK	11-05-2012 20:20:31	54d 7h 50m 52s	1/3	PING OK - Packet loss = 0%, RTA = 2.00 ms
ap-wrt-54gl-k2	PING	OK	11-05-2012 20:23:32	0d 9h 26m 30s	1/3	PING OK - Packet loss = 0%, RTA = 0.66 ms

Fig. 3. View Service Status Details for All Host Groups

Host	% Time Up	% Time Down	% Time Unreachable	% Time Undetermined
jupiter.vtu.bg	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
mail.vtu.bg	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
nas.vtu.bg	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
svn.vtu.bg	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
venus.vtu.bg	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
www.vtu.bg	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
Average	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%

Fig. 4. Host State Breakdowns

So far, the system monitors about 150 units and 200 information services, and this load utilize only a few percent of the server CPU capacity with very limited load kBps on the network interfaces as shown in Fig. 5, 6.

```
top - 21:20:58 up 19 days, 13:34,  1 user,  load average: 0.06, 0.03, 0.01
Tasks: 112 total,   1 running, 111 sleeping,   0 stopped,   0 zombie
Cpu(s):  1.7%us,  1.3%sy,  0.0%ni, 96.5%id,  0.2%wa,  0.2%hi,  0.2%si,  0.0%st
Mem:   1034096k total,   997748k used,    36348k free,   175328k buffers
Swap:  2097144k total,        0k used,  2097144k free,   305944k cached

  PID USER      PR  NI  VIRT  RES  SHR S %CPU %MEM    TIME+  COMMAND
23620 nagios    15   0 17672 2180 1176 S  0.7  0.2  16:41.49 nagios
27833 tato      15   0  2432 1076  804 R  0.3  0.1   0:02.02 top
    1 root      15   0  2168  588  512 S  0.0  0.1   3:34.93 init
    2 root      RT  -5     0    0    0 S  0.0  0.0   0:00.00 migration/0
```

Fig. 5. CPU (Memory) load measurements

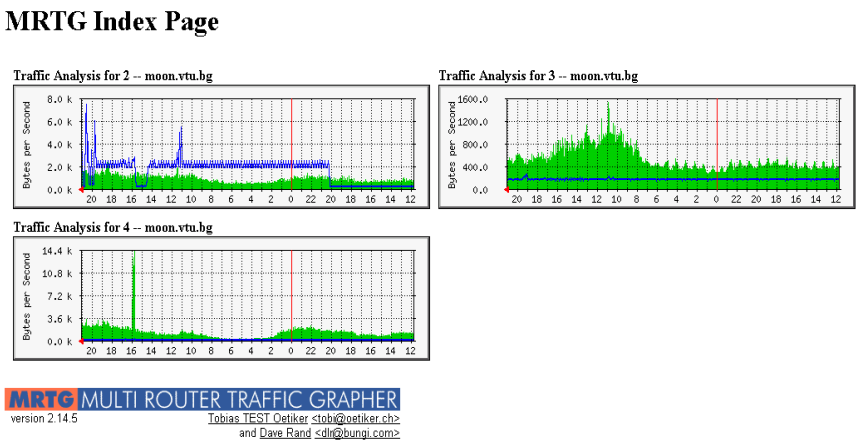


Fig. 6. Network traffic interfaces measurement

4 Conclusion

This network was developed as part of a University project aiming to provide high quality internet access and network services on the campus of ”Todor Kableshkov” University of Transport.

The development of software settings and distribution location of hardware devices of the network led to decrease in the number of access disruptions and network crashes, and ensuring of optimal operation and system load.

The establishment of a monitoring system for operation of the system lead to improved service quality and better trouble shooting. The implementation of this project led to accumulation significant expert knowledge as well as practical experience in designing, building and maintenance of complex computer networks.

References

1. Dimitrov D., Network logical development in university campus, Proceedings of International Conference on Application of Information and Communication Technology in Economy and Education – ICAICTEE-2011, p. 372-377.
2. Technical documentation, structure, patterns established communication and information network systems Todor Kableshkov University of Transport.
3. http://www.nagios.org/

Utility Company Consumers Consumption Patterns related to Illegal Customer behaviour (Thefts)

Andrey Bachvarov[1], Petko Ruskov[2], Kaloyan Haralampiev[2], Stefan Apostolov[3]

[1]Balkans Investment Consulting Agency Ltd.,andrey.bachvarov@bica-bg.com
[2]Sofia University,petkor@fmi.uni-sofia.bg; k_haralampiev@hotmail.com
[3]CEZ Bulgaria

"Historical data does become critical in decision-making when it is combined with current data and continuously refreshed with new information. This approach is the basis of predictive decision-making; models are based on what has happened and why, and also on what is happening now and what will likely happen next."
[IBM predictive analytics]

Abstract: Non-technical losses (NTL), more commonly known as thefts or illegal consumption of electrical energy pose an increasingly important problem for electricity distribution companies. The latter are facing an ever tightening legislative control and continuously decreasing margins so the importance of the reduction and prevention of the above losses has become a matter of survival for most energy companies.

Our goal is to learn how to identify the patterns indicative of illegal behaviour for the purpose of early detection, prevention and discouragement of similar conduct. To achieve this ambitious goal, this paper combines the results of academic research of applicable best practices, empirical tests based on real data obtained with the kind permission of a local distribution company and the analytical results obtained through the use of advanced statistical software. Our research greatly relies on data provided by the Operations department of "CEZ Electro Bulgaria" AD and subsequent sophisticated data mining analysis in order to detect and expose the consumers involved with electricity theft. We further investigate how those patterns of consumption can be modelled to show customer's risk behaviour for the specific needs of utility companies. The results include electricity consumption patterns developed through the use of the CRISP-DM research methodology using IBM SPSS Modeller predictive models and tools. We believe that the generated models, patterns and results can be readily applied by utility companies, their partners and consultants for future energy efficiency processes innovation and business transformation.

Key words: Electricity theft, electricity consumption, customer's risk behaviour, predictive model analysis

Introduction

Company information systems, sales transactions statistics, Web logs, social data records, Internet search indexing, call centres call and chat records, e-commerce, macro-economic data, competitive data and so on are just a tiny representative small part of collected data in companies nowadays. The volume of business data worldwide across all companies doubles every 1.2 years, urging the executives to take into account that overwhelming growth of data and developing the proper tools to handle what the enterprise really needs to be prepared for [22]. Noting that the above mentioned information is predominantly unstructured data, the necessity for analytical insights is gradually transforming itself from "good to have" to "must have" [13, 16, 19] and these, being put into due of data and context, will hopefully begin to reveal insights that can be realistically applied across the value chain in the whole organization and its partners. The idea that data and analytics can be employed to build competitive advantages and serve as the engine for innovation is neither new nor alien to the utilities companies [1, 4, 20, 23]. Research and implementation of energy efficiency processes allow utility companies to use less resources (including energy) to provide the same service to the customers [6, 7, 9]. End-user consumptions, especially in the residential sector is also attractive analysis for managers [14, 15]

Alternatively an innovative business model can either create a new market or allow a company to create and exploit new opportunities in existing markets [5, 12, 21]. Business model innovation takes place in a number of ways [3]:

- By adding innovative activities;
- By linking activities in innovative ways;
- By changing one or more parties that perform any of the activities;

The goal of this paper is to provide an empirical research about the home electricity consumptions and to discover theft patterns of electricity consumption and customer's risk behaviour. We investigate how patterns of consumption in utility companies can be modelled to show customer's risk behaviour.

2. Methodology and Research Design

We have applied a combination of analysis of a sample data base with real customer records, provided by Operations department of "CEZ Electro Bulgaria" AD, sophisticated analytical skills and data mining in order to detect and highlight the people responsible for electricity theft. The methodology is based on a CRISP-DM research methodology using IBM SPSS Modeller predictive models and tools – fig. 1. [10, 11, 18].

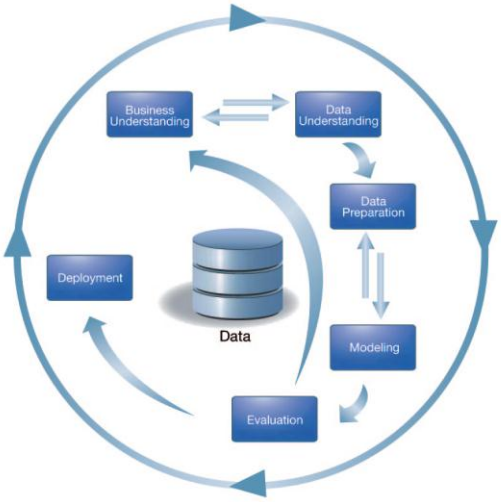


Fig. 1. CRISP – DM methodology

We have applied data mining techniques using IBM SPSS Modeller. More specifically we used the following tools:

- Auto Classify Node, which allows to use a qualitative (nominal or ordinal) variable as target;
- Auto Numeric Node, which allows to use a quantitative (scale) variable as target;
- Anomaly Node, which allows to identify costumers with atypical consumption patterns.

3. Data and Empirical Research

The next step of the research was the preparation of data set and models design.

3.1. Data

Our sample is drawn from a database provided by the Operations department of "CEZ Electro Bulgaria" AD [8], from what we believe is sufficiently large data set of customer records, combined with approximately 850 confirmed cases involving illegal electricity consumption and theft protocols for a period of 16 months from January 2011 to April 2012. The consumption is divided in three different tariffs – day, night and complete.

3.2. Models

We have built nine models, which are presented schematically in Figure 2. There are nine models for each of the abovementioned three tools (Auto Classify, Auto Numeric and Anomaly) we have three models corresponding to the three tariffs (day, night, and complete). These models are presented in the upper, lower and right sides of Figure 2.

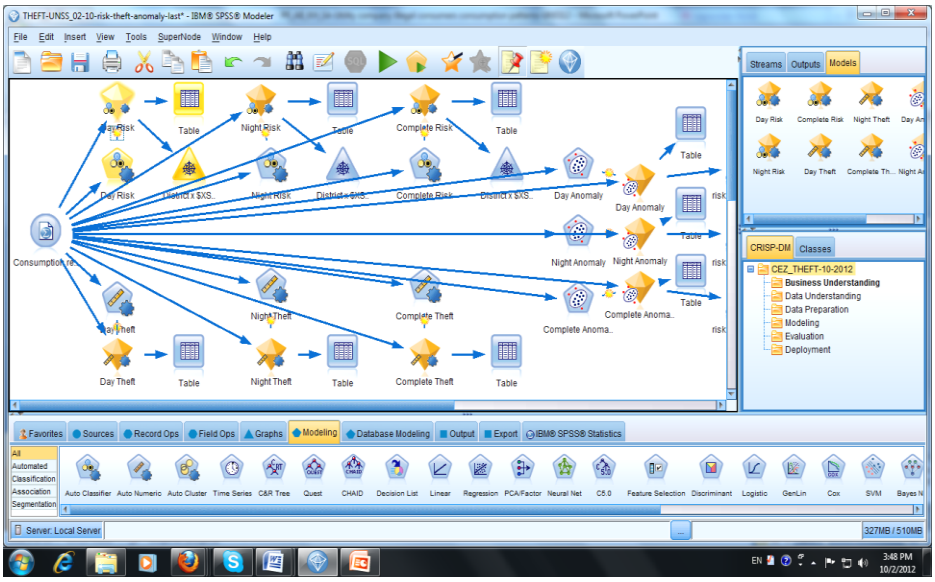


Fig. 2. Entire Model including Auto Classify Nodes (top), Auto Numeric Nodes (bottom) and Anomaly Node (right).

4. Discussions and Value for utility company

According to our methodology we have created nine models - 3 models with Auto Classify, 3 models with Auto Numeric and 3 models with Anomaly.

4.1. Auto Classify Node

To illustrate the Auto Classify Node we apply the data for daily electricity consumption tariff. The target variable, which indicates the risk of electricity theft, has three categories that were coded as follows – Table 1.

Table 1. Target variables code

Code	***Target variable***
0	No thefts detected;
1	Thefts detected;
2	Not inspected.

For research purposes we considered only the thefts in April 2012, which is the last month where inspection protocols exist.

The predictors, which we used to build an early warning model, which directs our attention to the potential risk users, was the residential area (Postal Code), district and monthly electricity consumption since January 2011 to March 2012.

Auto Classify Node allows us to experiment with all models in which the target variable is qualitative (nominal or ordinal), whatever predictors are. Supported model types

include C5, Logistic Regression, Bayesian Networks, Discriminant Analysis, KNN Algorithm, SVM, C&R Tree, QUEST, CHAID and Neural Net.

As a result of the application of the Auto Classify Node we get three best models – C5, CHAID and C & R Tree – Fig. 3.

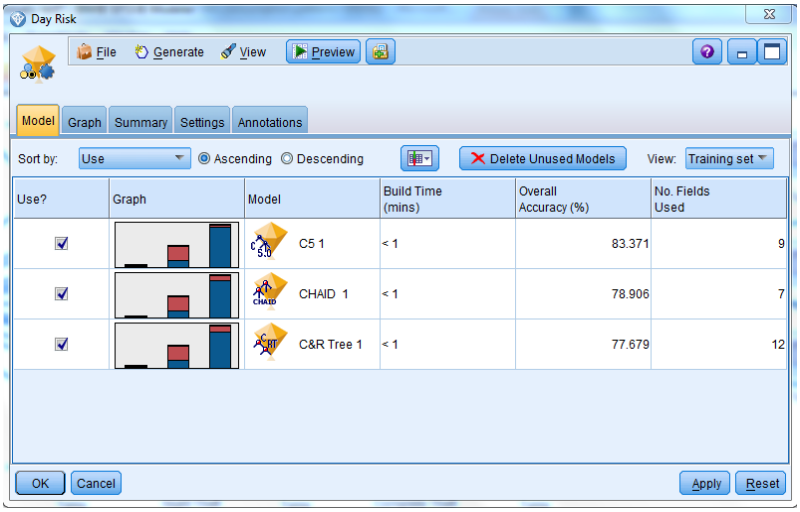


Fig. 3. The three best models obtained by Auto Classify Node.

These three models are versions of the so-called Classification Trees, so we will examine in detail only the first (the best) tree.

But first we started with the results, which are common to the three best methods. These are two charts – Fig. 4 and 5.

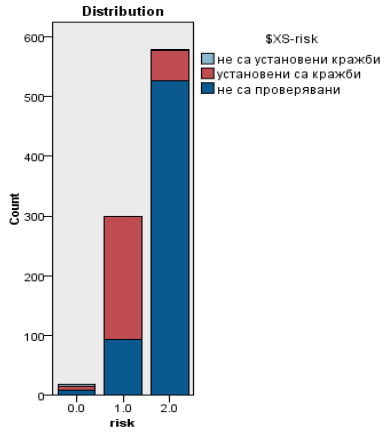


Fig. 4. Correctly and incorrectly classified users

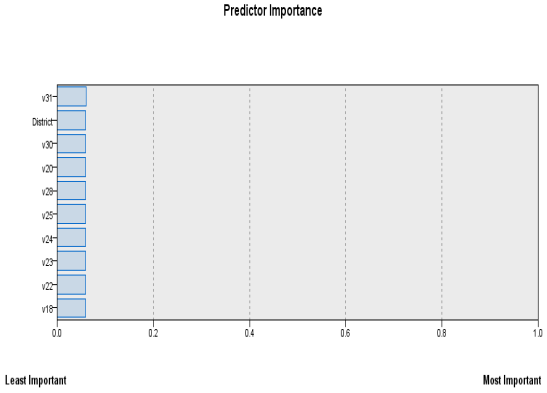


Fig. 5. Predictor importance

Figure 4 shows that about two-thirds of consumers with detected thefts are classified correctly, and the majority of consumers who have not yet inspected also are classified correctly. Of particular interest to us is that little part of not inspected users is classified as customers with detected thefts. These are users who are identified as risky by the "early warning system". They could be recommended to CEZ for future inspection. SPSS Modeler allows Table Node to be added to results, highlighting groups of suspicious users, recommended for further inspection.

Figure 5 shows the most important predictors. It can be seen that the most important predictor is consumption in February 2012. There are two alternative hypotheses:

1. For the "early warning system" is very important what the consumption in February was;
2. Since February is two months away from April(month during which the thefts are detected), for the "early warning system" is very important what the consumption was two months before the month in which we inspect for thefts.

Verification of these hypotheses is the subject of future work.

Since this Classification Tree is very large, we show only a fragment of it – Fig. 6.

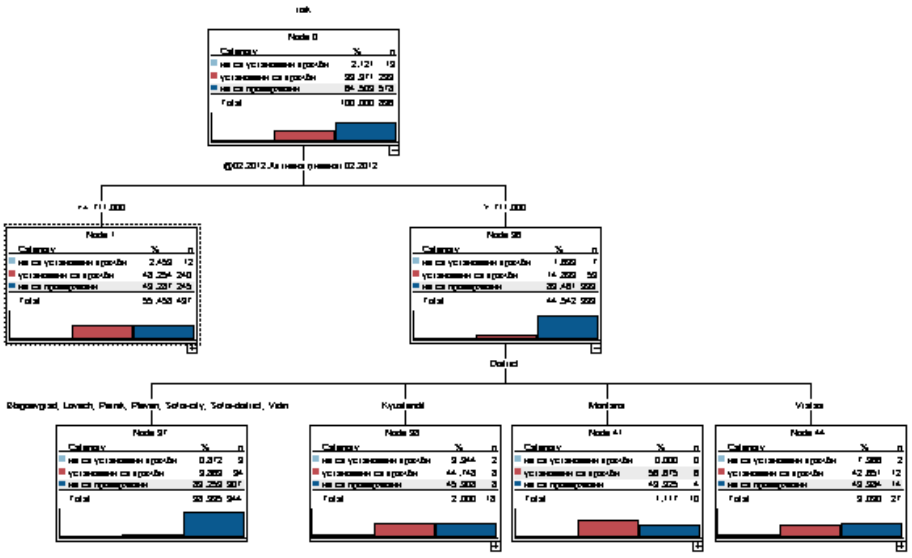


Fig. 6. C5 Classification Tree (fragment).

Thefts are detected in almost half of the examined instances consumers, according to the input data, described in 3.1, whose daily consumption during the month of February 2012 was under 711 KWh. For the users with a daily consumption of more than 711 KWh, the percentage of detected thefts is very low, but there are significant differences by districts. In the districts of Montana, Kyustendil and Vratsa, the theft rate is much higher than in other districts of the country. This analysis could be extended by unfolding of the tree and the successive study of all of its nodes.

4.2. Auto Numeric Node

To illustrate Auto Numeric Node we used only the data for night consumption. The target variable, which indicates the risk of electricity theft, is so-called additionally charged energy, i.e. amount of stolen electricity.

The predictors, which we use to build an early warning model, which directs our attention to the potential risk users, are the same as in the Auto Classify Node.

Auto Numeric Node allows us to experiment all models in which the target variable is quantitative (scale), whatever predictors are. Supported model types include Regression, Generalized Linear, KNN Algorithm, SVM, C&R Tree, CHAID, Neural Net and Linear.

As a result of the application of Auto Numeric Node we get three best models – KNN Algorithm, CHAID and Generalized Linear – Fig. 7.

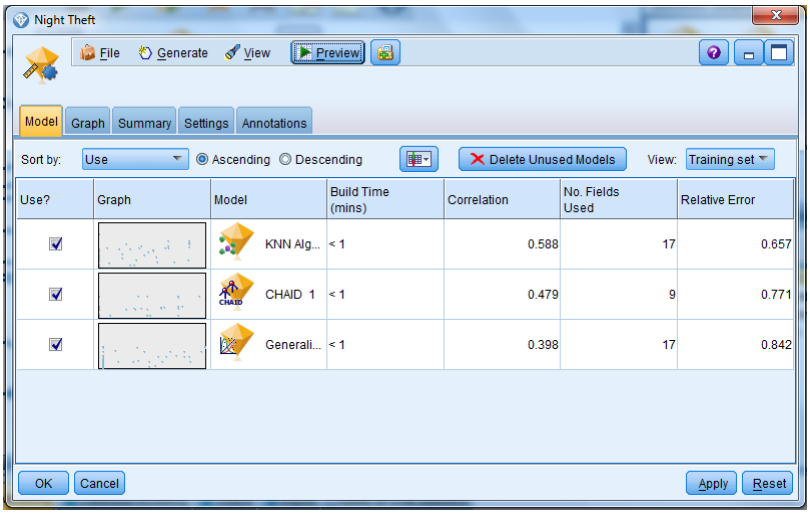


Fig. 7. The three best models obtained by Auto Numeric Node.

Again we started with results that are common to the three best methods. This is a graph with predictor importance – Fig. 8.

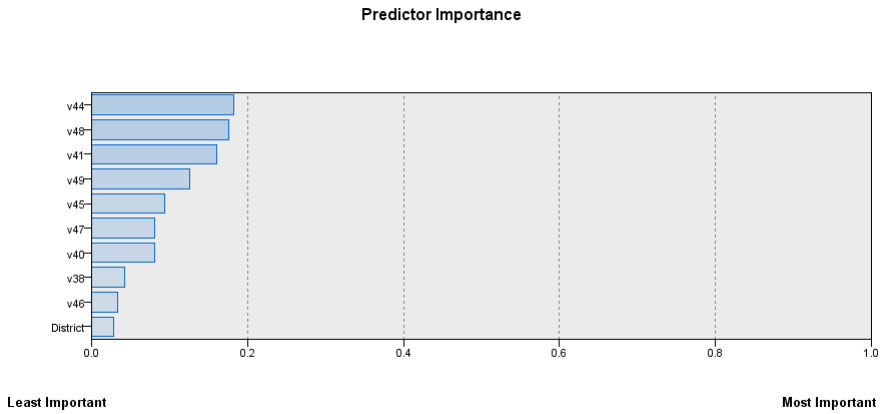


Fig. 8. Predictor importance

Figure 8 shows the most important predictors for potential risk users. It can be seen that the most important predictor is the energy consumption in October 2011, followed by consumption in February 2012 and July 2011. Again, various hypotheses, similar to the previous case, are possible. The definition of these hypotheses and their verification is a subject of future work.

Since in this case the three best models are independent, and they are not versions of a common model, here we only summarize the results instead of showing particular tables and graphs.

The relationship between night consumption in October 2011 and the additionally charged energy in April 2012 is a straight correlation–the increase in the night consumption leads to increase in the energy charged additionally, i.e. in most instances we are discussing stolen energy. This creates incentives for additional thefts in instances of high night consumption during this month. The relationship between night consumption in July 2011 and February 2012 and the additionally charged energy in April 2012 is opposite–if the night consumption in these two months is bigger, then the additionally charged energy is less, i.e. the low night consumption in these two months is risky.

There is a repetitive geographical pattern related to users from the districts of Montana, Kyustendil and Vratsa being most risky in the above context.

4.3. Anomaly Node

To illustrate Anomaly Node results we used only the data for complete consumption tariff. In this case there is no target variable, which indicates the risk of electricity theft, and all predictors are equal. To build an early warning model, which directs our attention to the potential risk users, we use only monthly electricity consumption since January 2011 to March 2012.

As a result of the application of Anomaly Node we get users with atypical complete consumption of electricity. The results of Anomaly Node can be related to actual detected thefts as follows:

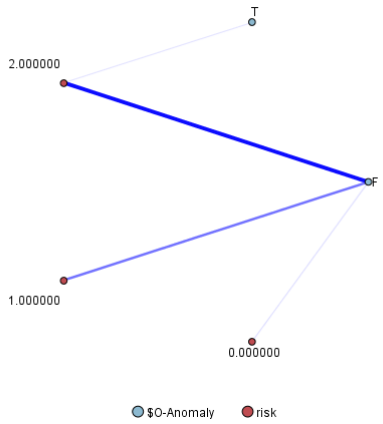


Fig. 9. Relationship between atypical consumption and electricity thefts

From figure 9 impress two things. First, all users with detected thefts are associated with F (False), i.e. all users with detected thefts have typical complete consumption. All users with atypical complete consumption (T – True) are not inspected. To identify them, to the results could be added a Table Node, where could be seen groups in which the users are classified. The users with atypical consumption could be recommended for inspection.

5. Conclusionand future work

In this research we have investigated how patterns of consumption in utility companies can be modelled and how to show customer's risk and theft behaviour.The results include electricity consumption patterns based on a CRISP-DM research methodology using IBM SPSS Modeller predictive models and tools.

The generated models, crucial patterns recognition and results can be applied by utility companies, their partners and consultants for future energy efficiency processes innovation and business models transformation.

Our future work plans are:

- To integrate into the models other input variables such as statistical climate data, demographic data, specific social and economic parameters and to discover more inside patterns;
- To reengineer and improve the mechanism aimed to identify risky consumers and forecast prevention of thefts based on identified patterns;
- To model as-is customer management process and do build to-be processes;
- To innovate and create effective business processes of the utility company;
- To use business analytics for new innovative business model transformation.

Acknowledgement

The authors are grateful to the Marketing & Sales support department of "CEZ Electro Bulgaria" AD for the explicit sample data base and helpful information.

The work on this paper has been sponsored by IBS Bulgaria, IBM Premier Business partner

References

1. A-L Lindén and A Carlsson-Kanyama, Energy Efficiency in Residences - A challenge for women and men, Energy Policy, 35, 2007, pp. 2163-2172.
2. Ambuj D. Sagara, Bob van der Zwaana, Technological innovation in the energy sector: R&D, deployment, and learning-by-doing, Energy Policy 34 (2006) 2601–2608.
3. Amit R., Christoph Zott, Creating Value Through Business Model Innovation, MIT SLOAN MANAGEMENT REVIEW, Spring 2012, Reprint 53310, Vol. 53, No 3, pages 41-49.
4. Bachvarov A., Petko Ruskov, Kaloyan Haralampiev, Electricity end users' consumption patterns in different west Bulgarian locations, UNITECH'12 International scientific conference, 16 – 17 November 2012, Gabrovo.
5. Baldwin E., Innovation Isn't 'Creativity,'It's a Discipline You Manage, MIT SLOAN MANAGEMENT REVIEW, April 2010.
6. Bojda, Nicholas, Business Case for Energy Efficiency in Support of Climate Change Mitigation, Economic and Societal Benefits in the United States, Lawrence Berkeley National Laboratory, 09-07-2011, http://escholarship.org/uc/item/7xm4m8ht
7. Carl Haigney, Customer Relationships and the Economy, Shaping a New Era in Energy, www.UtilitiesProject.com.
8. CEZ Electro Bulgaria AD http://www.cez.bg/en/home.html, http://www.cez.bg/en/customer-service/ways-of-payment.html.
9. Druckman A., T. Jackson, Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model, Energy Policy, - Elsevier, 2008.
10. IBM Corp., CRISP-DM 1.0, Step-by-step data mining guide, © Copyright IBM Corporation 2010.

11. IBM Corp., Switching perspectives, creating new business models for a changing world of energy, 2010.
12. Kiron D., Pamela Kirk Prentice, Renee Boucher Ferguson, Innovating With Analytics, MIT SLOAN MANAGEMENT REVIEW, VOL.54 NO. 1, 2012
13. LaValle S., Michael Hopkins, Eric Lesser, Rebecca Shockley and Nina Kruschwitz, Analytics: The new path to value, How the smartest organizations are embedding analytics to transform insights into action, IBM Institute for Business Value, MIT Sloan Management Review, 2010.
14. Letschert V., The Boom of Electricity Demand in the Residential Sector in the Developing World and the Potential for Energy Efficiency, Lawrence Berkeley National Laboratory, 02-05-2010.
15. Lukas G. Swan, V. Ismet Ugursal, Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, Renewable and Sustainable Energy Reviews 13 (2009) 1819–1835, 2009.
16. Marc Teerlink and Michael Haydock, Customer analytics pay off, Driving top-line growth by bringing science to the art of marketing, Copyright IBM Corporation 2011.
17. Miles I., Patterns of innovation in service industries, IBM Systems journal, vol.47, No 1, 2008, pp.115-128.
18. Morelli T., Shearer C., Buecker A., IBM SPSS predictive analytics: Optimizing decisions at the point of impact, IBM redpaper 4710, Copyright IBM Corp. 2010.
19. Pearson N., Eric Lesser and Joel Sapp, IBM Global Business Services, Executive Report, IBM Institute for Business Value, A new way of working, Insights from global leaders, 2010.
20. Peter Skarzynski, Rowan Gibson, Innovation to the Core A Blueprint for Transforming the Way Your Company Innovates, Harvard Business School Publishing Corporation, 2008.
21. Ruskov P., A. Bachvarov, Disruptive innovations for the energy system, Anniversary Scientific Conference with International Participation "40 Years Department of Industrial Automation", 18 March 2011, Sofia, Bulgaria, pp. 141-144.
22. Rohit Tandon R., Arnab Chakraborty and Ganga Ganapathi, Leveraging Big Data around customers, Analytics, September/October 2012, http://www.analytics-magazine.org/september-october-2012/659-marketing-leveraging-big-data-around-customersoctober-2012/659-marketing-leveraging-big-data-around-customers.
23. Timothy J. Foxona, Geoffrey P. Hammondb and Peter J.G. (2010), Developing transition pathways for a low carbon electricity system in the UK, Pearsonc Technological Forecasting and Social Change, Issue includes a Special Section on "Infrastructures and Transitions", Volume 77, Issue 8, October 2010, Elsevier Inc., pp. 1203-1213.

Advantages of measurement systems based on digital sensors, estimating the real effect of an energy saving project application

Nanko Bozukov, Hristo Dinkov, Ivan Gaytandzhiev, Marko Dimitrov

Abstract. In the paper the advantages given of DS1820 temperature sensors in building measurement systems for different purposes, especially for the control of the effects after energy saving project for buildings has been applied are pointed out. Decisions for creating a simple structure reliable system are shown. Conclusions about reliability of the system under different conditions concerning the number of sensors, length of the bus and cable used have been drawn. Some advantages which concerns the direct addressing possibility of the sensors, connected in the network are pointed out.

The modern information and control systems collect different type of sensor information, which is processed and stored in a various type format. In the paper the advantages given by 1-wire digital sensors for temperature, relative humidity and so on are applied in that kind of systems are pointed out. These sensors are extremely useful for creating systems for temperature control and measurement, with a simple architecture and a large number of measurement points, because of their own ability to sample, convert and transmit data. Their advantages are connected not only with their low cost, but also with the accuracy and simplicity when working with them. In fact, in some cases it may occur problems with the reliability of the system concerning transmission and reception of the data. Practically, this is the key indicator for the sustainability of data collecting and saving systems. In this meaning, we should be very careful when making our choice for the communication adapter we are going to use.

The presence on the market of a temperature sensors as a DS18S20 from Dallas Semiconductor significantly simplify the task of creation and when necessary after that, an easy expansion of the number of measurement points [1,2,3]. The most important characteristic for this kind of sensors are given below:

- Linear characteristic across the range;
- Own unique address for communication protocol 1-Wire Bus, specially developed by Dallas Semiconductor;
- Transmission of the data in digital form;
- Suitable temperature range for habitable rooms and outdoor environments, from -55 to +125 0C.

As it was mention above to avoid communication problems we must carefully choose a suitable communication adapter for any particular case of application. In fact it is easy to do that, because of a serious support of different standard devises such as – parallel or serial buss adapters and these specially designed for a long distance communication one.

The particular choice depends mainly on the topology of the system- the length of the buss and the number of the connected sensors.

In our case we propose the following most simple structure of the measurement system:

1. PC
2. Dallas Semiconductor serial port communication adapter
3. Dallas Semiconductor temperature sensors - DS 18S20

Developing PC based 1- Wire buss systems, it is very important to have in mind certain restrictions due to the chosen topology [4, 5, 6].

Figure 1 shows three possible types of topologies:

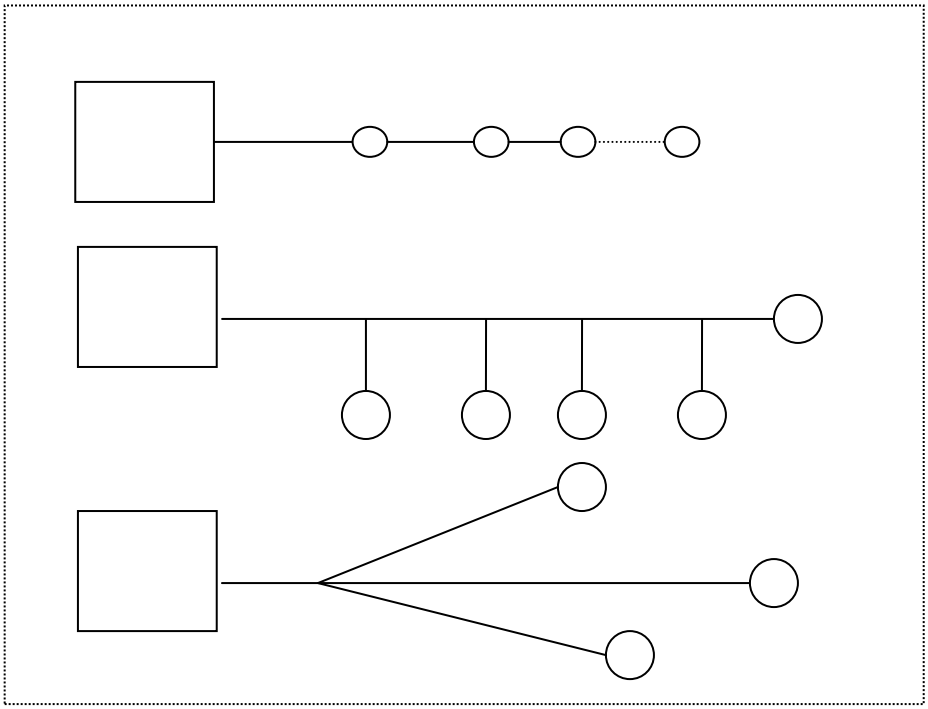


Fig. 1. Topology of systems with PC and 1- Wire Buss

1. Linear – when distance between the buss and sensors are minimal.
2. Tree- type – when the connection between the main buss and sensors are with a different length.
3. Star type – long individual branches and a minimal main buss.

Because the sensors are parasitic power supplied (from the buss), as a rule, the total cable length of the main buss and the branches should not exceed a determined of the manufacturer of the chosen communication adapter. It is strongly recommended branch length not to exceed 3 meter. Usually that leads to the use of the first type of the topologies pointed out above (except the cases, when the sensors are situated near to the adapter and it is possible to use the third one).

PC based system with a suitable UPS gives an opportunity not only to collect and store measurement data but they could be easily processed for different purposes. Another

advantage in using 1-Wire buss components is that the software developed from the manufacturers is aimed to collect and record data in text file format and most commonly in tree structure- in days, mounts and years. The text file could be easily imported in standard and widespread software package such as MS EXCEL, and the data could be easily analysed, processed and as a result displayed graphically.

By means of user designed application software module, different tasks could be additionally performed:

1. Dynamic table generation for all sensors connected to the system.
2. A daily, weekly or other time interval report- either in table or graphical format.
3. Document printing.
4. Database keeping.
5. Specific alarm defining and generation.
6. On-line calculation for different technological parameters.

Other significant advantage of the systems like this is that they have an open structure about the number of sensors, respectively the control points. This approach allows extending the system any time, without the need of software or hardware changes and almost without additional expenses. It is important, that the sensors have their own unique address, and software can automatically determine their presence in the net and to define the correct position of the controlled point.

In the field of Energy efficiency control, in the meaning of Energy Efficiency Program and European strategy for sustainable and competitive energy, those systems, because of their simplicity and low cost[7]. They also could be easily applicable and very useful in post implementation control, especially when estimating the effect of the undertaken project measures and their real values. All accumulated data about room temperatures, outside temperatures, the average temperature inside for the period, energy consumption for the period and other values, gives an objective background for parameters calculation and the comparison between project and real values. This is an objective base to make the final conclusion, if the undertaken project measures had reached the project demands for energy saving and lowering the costs, or they had not. In this case the energy efficiency certificate for the building will be based on a database of real measured data for the passed period of time.

Conclusion

• Information and data logging system is a very useful and effective tool for energy saving post implementation control and the objective effect estimation.

• The collected and stored data can be used directly by the officially recognized systems for energy flows calculation, because of their text form format and easy transfer.

• The exceptionally low cost and simplicity of the system, especially relative to the overall project investment amount leads to the idea, this type of measurement system to be included just at the beginning phase of the project and as a first phase during implementation.

• Overall on-line temperature monitoring in- and outside building, contributes and supports operative and in-time control action from operator in order to minimize power waist and this way to optimize energy expenses.

References

[1]. Dallas Semiconductor Corp. DS1820: 1-WireTM Digital Thermometer, Dallas, Texas, 1998, 1 – 27.
[2]. Dallas Semiconductor Corp., Book of DS19xx iButton Standards, Dallas, Texas, 1995, pp. 35÷64.
[3]. Dallas Semiconductor Corp., Automatic Identification Data Book, Dallas, Texas, 1995, pp. 117÷136.
[4]. Roseman M., Meta Kit: Quick and Easy Storage for your Tcl Application, Equi4 Software - Draft, 2002, http://www.equi4.com/metakit/tcl.html
[5]. http://www.maxim-ic.com
[6]. http://www.vektor1.bg
[7]. Николова Н., М.Стоев, Европейската стратегия за устойчива и конкурентна енергетика и мястото на България в Европейската енергийна система, Международен журнал „Устойчиво развитие", брой 1, юли 2011, Издателство на Технически университет – Варна, с.63-69 (ISSN 1314-4138)

DATA AND METHODS FOR MEASUREMENT OF ENERGY SAVINGS

Nanko Bozukov[1], Marko Dimitrov[2], Hristo Dinkov[1], Ivan Gaitandjiev[3]

[1]UFT- Plovdiv, bozukovnanko@abv.bg, [2]"KEO-Zenith", Ltd, Sofia, dimitrovmarko@abv.bg,
[1]UFT- Plovdiv, hristo.d @ dir.bg,[3]"Vector-1" Ltd., Plovdiv, vektor1@infotel.bg

Abstract. The report examined data and program data processing, which is essential for precise and accurate assessment of energy savings in developed and validated methods. Data and methods are based on measurements or estimates. It is important that before any assessment accurate information and a set of necessary data can be provided. In assessing the energy service or measure energy efficiency, usually can not rely on specific measurements. It is therefore necessary to distinguish between methods measuring energy savings and methods estimating energy savings, the latter is more common and are associated with much lower costs.Developed methodologies for assessing the energy savings. In the development of each is respected principles for enabling the evaluation of energy savings is achieved by measuring and / or estimating consumption before and after application of the measure for energy efficiency.

Key words: methods measuring and estimating energy savings, energy efficiency

Introduction

In developing methods for assessing the energy savings freedoms principle is observed by measuring and / or estimating consumption before and after the implementation of the energy efficiency measure. Developed methodologies cover the following areas:

- Replacement of equipment with energy efficient;
- Updating equipment or building;
- Acquisition of energy efficient equipment or building.

The amount of the saved energy is the difference between the used energy before and after the introduction of the measure or program to improve energy efficiency [6].

Block diagram

The algorithm of the program for data processing is described in block diagram shown in fig. 1. The program was developed in the programming language of the scientific and engineering calculations system MatLab [2, 3, 4], which is a dialog system and program for scientific and engineering calculations and visualization based on matrix representation of the data.

MATLAB is open, flexible and easily adaptable to the decision of a user's task. It is applicable for many fields of knowledge, because it contains tools, including data processing, operations on matrices, vectors, polynomials, complex numbers, signal analysis, digital filtering, two-dimensional and three-dimensional graphics and more.

Description of the block diagram of the data processing program.

After starting the program enter the number (1, 2, 3, ... 12) of the used assessing energy savings methodology. Each method corresponds to a subroutine that is executed and the results are displayed on the operation of the subroutine. If necessary, perform another subroutine. After completion of the work all the results are displayed.

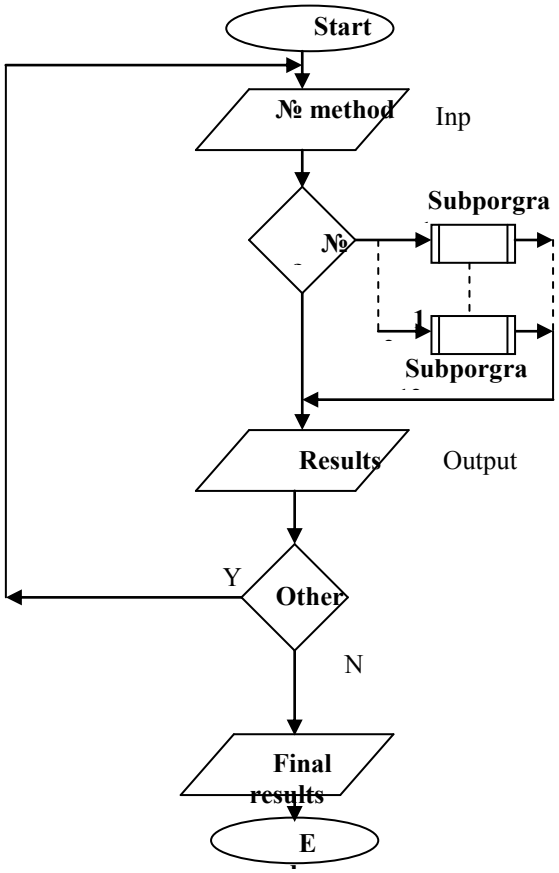


Fig. 1. Block - diagram of the program for data processing

All variables used in the program, and the entire session working with the program are stored in a separate file that can be opened, viewed and printed.

Benefits of the data processing program

The program is easy to process data necessary to implement methodologies that aim to demonstrate the implementation of individual measures or groups of measures to improve energy efficiency in residential and public buildings.

The above program is useful in the post-implementation of energy efficiency measures. Based the accumulated database can more accurately assess the real impact of the measures – the maintained actual temperature in the building in terms of design temperature and depending on the outside temperature, fuel coolant, etc. On this basis, a complete analysis can be done and the planned project burning characteristics and achievements in the actual operation can be compared. The certificate issued within the period provided by the energy efficiency law would have a solid justification of factual data.

Conclusion

The system for data processing is a good and convenient means of control and objective analysis of the impact of the application of energy efficiency measures. Applying such a system can achieve high reliability, easy transfer and data processing. This data is used to calculate the system energy flows.

The availability of data and the possibility of collecting and processing them is essential for accurate and fair assessment of the energy savings in the developed methodologies. That's why it is important the most accurate information and a set of necessary data to be provided before any assessment.

References:

1. http://www1.bpo.bg/images/stories/buletini/binder-2012-09.pdf
2. VP Dyakonov, MATLAB 6: Training Course, St. Petersburg, Peter, 2001
3. Chen, C., P Dzhibling, A. Irving, MATLAB in mathematical research, Moscow, Mir, 2001
4. www.mathworks.com
5. http://www.eulaw.egov.bg/DocumentDisplay.aspx?ID=382929
6. http://www.mi.government.bg/files/useruploads/files/microsoft_word_-_metodiki.pdf

LIKELY-HOOD-BASED EDUCATIVE DATA ANALYSIS OF ALTERNATIVE TESTS WITH A SPHYGMOMANOMETER

Lubomir H. Lahtchev

Institute of Systems Engineering and Robotics, Bulgarian Academy of Sciences, e-mail: lahchev@icsr.bas.bg , lahchev@iser.bas.bg

Abstract*:* A brief overview of the comprehensive non-invasive sphygmomanometers shows the variety of locations for tests along the parts of the arm. The individual tests on other alternative parts of the arm provide complementary information about the blood pressure analysis and can reveal abundance of mutually dependent results. The sorted by medical ranges and test locations data for the blood pressure (BP) are quantity assessed on 12 aggregate arrays by 81 values. A comparison of both data from the recommended and from the alternative parts of the arm represents a new type of medical research. The statistic distributions commit descriptions of phenomena estimated by likely-hood- based educative data analysis, scientifically substantiated in the paper.

Key words: statistics, sphygmomanometer, assessment, data analysis, hypothesis.

INTRODUCTION

The wide disseminated method for measurement of the BP of Korotkov works at application on the upper arm.

Devices for non-invasive measurement of the BP from the upper arm, wrist and the finger are known. The upper arm still is preferred location of BP-measurement [1]. The estimations about measurements of the BP in the wrist have discrepancy and the finger is not recommended location.

Alternative tests of a sphygmomanometer in the current paper are individual measurements in the locations of the arm, which are different from the minded above. These are the forearm and the elbow, which can be real data sources for BP at cases, when the other locations are non-acceptable for measurements without change of the equipment – a sphygmomanometer for upper arm.

It is already known that the International Protocol [1] and the Standard EN 1060-4 [2] contain important disadvantages of a mathematical aspect, described in [3, 4], which substantiate the question about their mathematical basis. As a result, many sphygmomanometers can be found accurate and other inaccurate without defined double signed ranges of existence, non-adequacy mathematic provisions on the early stage of data processing. In this paper, double-signed domains of existence of the deviations are introduced and an algorithm for assessment of a hypothesis is suggested. The likely-hood-based educative data analysis is a foundation for inference. The current exploration is based on BP measurements, executed on an individual healthy human. The sphygmomanometer for the upper arm includes an aerobic cuff, which can also be located on the elbow and the forearm. Both the elbow and the forearm are locations for tests of the BP at alternative domains of the arm with a sphygmomanometer for the upper arm. These can provide BP data

for analysis extra, based on prepositions, logic expressions and having in mind results of their statistic processing. As a general, the purpose of the paper consists of comparisons and inferences, based on statistically assessed parameters: systolic, diastolic BP and statistic estimation of a hypothesis.

ASSUMPTIONS AND PRE-REQUISITES

The experimental measurements of the BP are completed with accordance to the recommendations of the producer [5] on a relaxed body of a human at the upper arms and also at two alternative points of the arms: elbows and forearms [6]. The experimental scheme of BP-measurements, represented in [6] is chosen in accordance to the principles of the singular micro- and macro-spaces: straight-linearity, parallelism, and balanced rotation.

Portions of measurements consist of 6 entities with a kept five-minute interval between each one pair of them for one hand. The typified domains of measurements of the BP are ordered by consequence of their chronology as follows: left forearm, right upper arm, left upper arm, right forearm, left arm elbow, right arm elbow. For convenience, the corresponding arrays of systolic and diastolic BP have been statistically processed by mathematical expectation, deviation of the random value from the mathematical expectation, dispersion and mean squared standard deviation described in [3, 4]. The author's presentation of this material is prolonged in [6] by statistic processing and estimation of the probability for $n_i = \alpha_i$ for each one aggregate array.

It is correctly accepted the random background of the BP data and each one array represents issues of a statistically ordered by chronology data series.

The physic co-ordinate of the BP with a measure unit millimeter of mercury [mmHg] usually covers the distance of 40 – 200 mmHg, including 40 – 110 and some extra for the diastolic BP, and 60 – 180 mmHg for the systolic BP. Both sub distances of diastolic and systolic BP are separated on four defined closed zones and one open zone for each one of them, incorporated from the medical practice, for example [1].

As whole, the distributions of twelve aggregate arrays by 81 issued values, called issues, for the left and right arms are classified by five zones of distribution, shown on the table 1.

Table 1. Data distribution of the issues by ranges of BP and locations of measurements.

Zones of the Systolic/Diastolic BP [mmHg]	Systolic / Diastolic BP on the right upper arm	Systolic / Diastolic BP on the left upper arm	Systolic / Diastolic BP on the right forearm	Systolic / Diastolic BP on the left forearm	Systolic / Diastolic BP on the right elbow	Systolic /Diastolic BP on the left elbow
60 - 100	**0**	**1**	**0**	**1**	**4**	**11**
101 - 130	**68**	**75**	**50**	**62**	**77**	**70**
131 - 160	**13**	**5**	**31**	**17**	**0**	**0**
161 - 180	**0**	**0**	**0**	**1**	**0**	**0**
>180	**0**	**0**	**0**	**0**	**0**	**0**
40 - 60	**0**	**1**	**0**	**1**	**3**	**3**
61 - 80	**49**	**57**	**19**	**31**	**51**	**53**
81 - 100	**32**	**23**	**59**	**46**	**27**	**24**
101 - 110	**0**	**0**	**2**	**3**	**0**	**1**
>110	**0**	**0**	**1**	**0**	**0**	**0**

LIKELY-HOOD-BASED SIMILARITY

A hypothesis, noted A, of similarity of several different random values of corresponding to them signals can be described by the expression:

"THE BLOOD PRESSURE WITHIN THE FOREARM EXCEEDS THE BLOOD PRESSURE WITHIN THE UPPER ARM."

The concise formulation here will easier discover the peculiarities of the following process of description, which can also serve for educative purposes.

The positive probability of a hypothesis to meet the requirement of its formulation is named *likely-hood of a hypothesis*, noted as $p_{\Sigma}(n)$. The negative probability of a hypothesis to be wrong, when it does not meet the requirement of its formulation is named *likely-hood error of the hypothesis*, noted as ε. Both are bound by the expression:

$$\varepsilon = 1 - p_{\Sigma}(n) \qquad (1)$$

The truth/non-truth about the hypothesis A here is established by algebraic difference between the BP-value, measured on forearm, minus that value, measured on upper arm.

The structure of aggregate area contains eight among the above twelve arrays bound to the hypothesis A. They are systolic and diastolic BP-data of the left and right upper arms and forearms. Four aggregate arrays belong to the upper arms, and other four – to the forearms. Thus, four or k=4 -comparisons will define the scores for analysis.

The general positive probability p_{gl} follows the k-variable, which varies within domain 0 – 1 as 0, ¼, 2/4, ¾, 4/4 and achieves the general number of inner general

probabilities K=5 for the currently assessed hypothesis A. Hence, the equation about the general positive probability is:

$$p_{gl} = k_l / K \quad (2)$$

where l – index of the k-value of the domain of existence.

The differential positive probability of the A-hypothesis is defined by relative frequency of positive comparisons of issues about forearms minus issues about upper arms for a current general positive probability p_{gl} or:

$$p_{di} = n_i / N \mid p_{gl} \quad (3)$$

n_i – number of positive individual comparisons at an accepted p_{gl}; N – number of general individual comparisons; N=81 for every one experimental group of p_{gl}.

The common likely-hood of A-hypothesis is computed by the equation:

$$p_{\sum}(n_{x,y}) = \sum_{i=0}^{K} p_{di}(n_{ix,y}) \cdot p_{gi} \quad (4)$$

The physical context of the common likely-hood $p_{\sum}(n_{x,y})$ represents integral – differential characteristics of the A-hypothesis for $N -> \infty$, $K -> \infty$. The corresponding likely-hood error is computed through eq. 1.

EXPERIMENTAL DATA DISTRIBUTION BY ALTERNATIVE PARTS OF THE ARMS

The experimental numerical distributions of the data about BP-measurements are shown as a pre-requisite above in the table 1.

The corresponding graphics on the fig. 1 is a base for data analysis. The vertical axis represents the values of data concentrations and the horizontal axis – the zones of BP-data. The light colored bricks belong to the upper arms and the forearms. These are closely concentrated. The dark colored bricks correspond to the elbows' data – concentrated for the systolic BP, but with a deviation toward low zone of BP and more dispersed toward both low and high zones of BP for the diastolic BP.

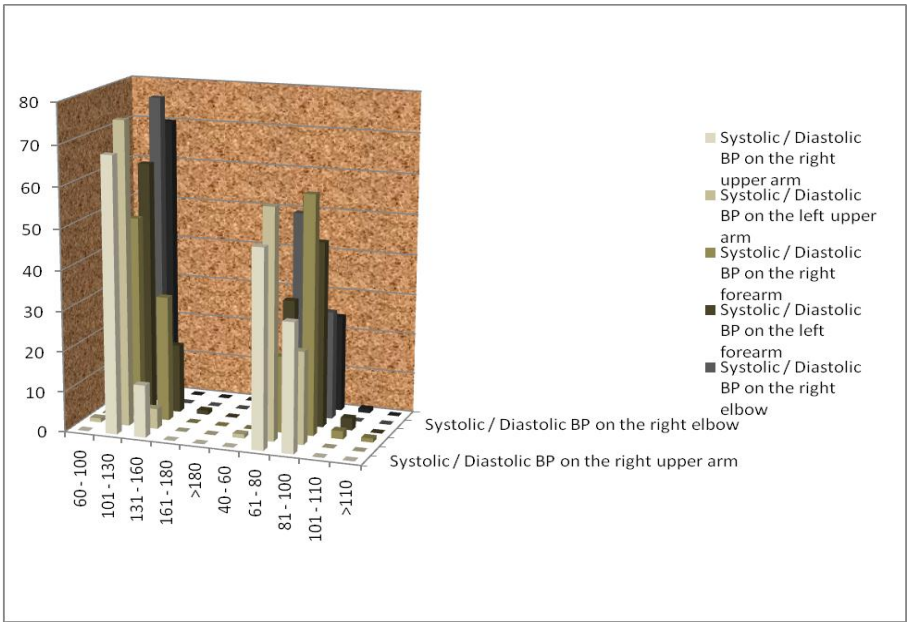


Figure 1. Data distribution by locations of measurements the BP.

Hence, several inferences on experimental data (IED) distributions can be formulated.

IED1. The measurement of the BP at the location of the upper arm contains data concentrated in two zones of normal and slightly over normal BP for the systolic BP. Similarly is the distribution of the diastolic BP. The difference between systolic and diastolic BP (see table 2) is closely to the normal and some over – 42-43 mmHg. The corresponding mean squared deviations on table 2 are lowest for systolic and diastolic BP and less than 3 mmHg. These results have not discrepancy. The signal is stable within the systolic zones and slightly dispersed within the diastolic zones.

IED2. The measurement of the BP at the location of the forearm contains dispersed within four BP-zones data, but still concentrated in two of them – normal and slightly over normal BP for the systolic BP. The values of systolic BP are equal and more than the corresponding values of systolic BP measured within the upper arm. The systolic and diastolic BP slightly and synchronously deviates to the higher BP-zones keeping similarity to the BP-signal within the upper arm. There is weakly expressed discrepancy of the distributions at normal difference between systolic and diastolic BP (see table 2) – 40-42 mmHg. The corresponding mean squared deviations on table 2 for systolic BP exceed the 3 mmHg boundary and it can be result either the oscillations or the higher sensitivity of the sphygmomanometer. Thus results for the forearm locations have some dispersion of the accuracy and weak discrepancy at the distributions, but still at clearly expressed similarity to the BP-signal within the upper arm.

IED3. The measurement of the BP at the locations of the elbow contain dispersed within four BP-zones data and wave form of the distributions slighting to lower systolic BP-zones and higher diastolic BP-zones, when the BP within two other locations strive to normal and slightly over normal BP. The systolic BP values descend to the lower BP-levels, when the diastolic BP – to some little bit under normal ones. At the same time, the difference

between systolic and diastolic BP (see table 2) keeps under normal – 35-36 mmHg. The corresponding mean squared deviations on table 2 are within the required boundary under 3 mmHg, but it could be result of a lower sensitivity and fussiness. Thus, we have "mutual attraction" of the systolic and diastolic BP-distributions, decreasing difference between systolic and diastolic BP-values, which prompt about fussiness of the BP-signal supporting its acceptable dispersion and strong non-sensitivity to the systolic BP-signal and small non-sensitivity to the diastolic BP-signal of the sphygmomanometer at this location and its results have important dispersions and discrepancies.

IED4. The similar data distributions of the systolic BP and diastolic BP of the upper arms and forearms and their average data on table 2 are a base for imply the hypothesis A: "The BP within the forearms exceeds the BP within the upper arms."

Table 2. The average values – mathematical expectations about BP and corresponding mean squared standard deviations.

		right arm	MSSD	left arm	MSSD
Upper arm	systolic BP	122.22	±1,336	119,43	±1,298
	diastolic BP	79.21	±1,1865	76,89	±1,175
Forearm	systolic BP	127,86	±1,7092	122,69	±1,5318
	diastolic BP	85,97	±1,321	82,037	±1,3049
Elbow	systolic BP	112,77	±1,2417	110,87	±1,3679
	diastolic BP	77,037	±1,3765	76,07	±1,4867

EXPERIMENTAL RESULTS BY LIKELY-HOOD-BASED ASSESSMENT OF A HYPOTHESIS

The A-hypothesis is proposed in the part of likely-hood-based similarity. In order to assess numerically its likely-hood the corresponding algorithm computes results of likely-hood – 0,7222 and the likely-hood error – 0,2778, shown on table 3. The end right column contains verification sums for the intermediate computations and final results at the lasts two lines.

Table 3. Likely-hood of a hypothesis.

General positive probability	0	0,25	0,5	0,75	1	
Data distribution by %	1,23457	9,87654	25,9259	24,6914	38,2716	100
Differential positive probability	0,01235	0,09877	0,25926	0,24691	0,3827	1
Hypothesis likely-hood	0	0,02469	0,12963	0,18519	0,38272	0,72222
Hypothesis likely-hood error						0,27778

The sorted by medical ranges and test locations data for the BP are quantity assessed on 8 aggregate arrays by 81 values for the upper arms and forearms. The data distribution on fig. 2 represents the practical results of the empirical algorithm suggested in this paper. Thus, the BP within the forearm exceeds the BP within the upper arm with an acceptable probability.

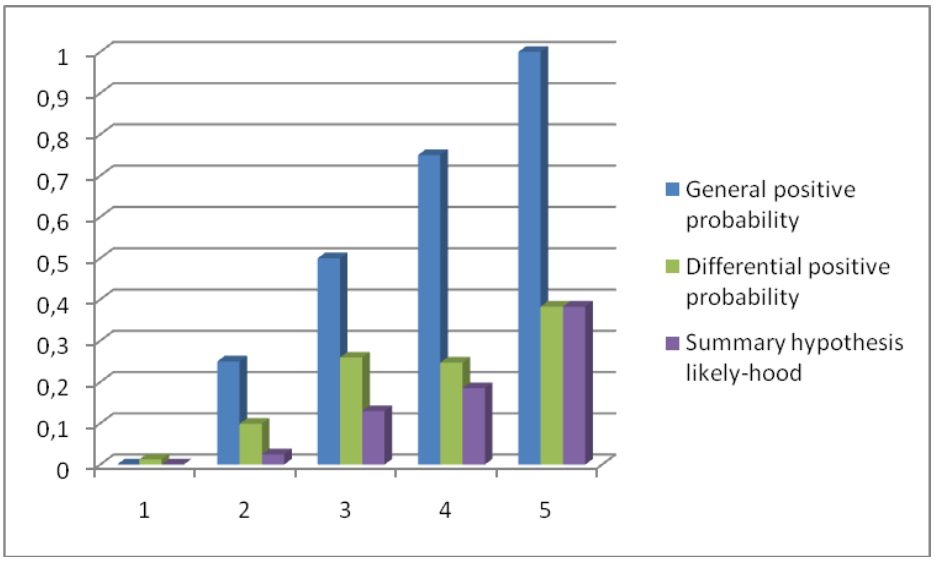


Figure 2. Data distribution at assessment of likely-hood of a hypothesis.

CONCLUSION

The individual tests on alternative parts of the arm with a non-invasive sphygmomanometer for the upper arm provide abundance of experimental data – 12 aggregate arrays by 81 numerical values – issues for assessment the measurements the BP. The data distributions by medical zones of BP and by locations of measurement are numerically and graphically represented. They serve at formulation the inferences with concern the signal about BP, the locations of measurements and the sphygmomanometer. A general positive probability and differential positive probability are introduced for assessment of a hypothesis and the corresponding algorithm of computation the likely-hood of a hypothesis is represented. The experimental results on this topic are also implied and demonstrate an example of educative data analysis. All data computations are verified on Excel.

AKNOWLEDGEMENT

The paper reflects some new characteristics and theoretical presumes through experiments on the explorations bound to the singularity of the micro- and macro-spaces. It also lasts the work started by the MI1509/2005.

REFERENCES

1. E. O'Brien; Th. Pickering; R. Asmar; M. Myers; G. Parati; J. Staessen; Th. Mengden; Y. Imai; B. Waeber; P. Palatini; W. Gerin; N. Atkins. *International Protocol for Validation of Blood Pressure Measuring Devices in Adults.* Blood Pressure Monitor., February 2002, vol.7, issue 1, pp. 3-17.
2. British Standard BS EN 1060-4: 2004. *Non-invasive Sphygmomanometers – Part 4: Test Procedures to Determine Overall Accuracy of Automated Non-invasive Sphygmomanometers.* 36 p.
3. L. Lahtchev. *Cloud Computing Individual Assessment of a Medical Device for the Education.* Proc. of Intern. Conf. on Appl. of Inf. and Com. Tech. in Ec. and Educ., December – 2011, UNWE, Sofia, pp. 503-512.
4. L. Lahtchev. *Individual Assessment of a Medical Device for the Education by Relative Features.* Proc. of Intern. Conf. on Appl. of Inf. and Com. Tech. in Ec. and Educ., Decem. – 2011, UNWE, Sofia, pp. 281-289.
5. Tensoval® Duo Control. *Instructions for use*. Published on the site: www.hartmann.info , www.tensoval.de 2008, pp. 126 – 147.
6. L. Lahtchev. *Cloud Computing Statistics of Individual Alternative Tests with a Sphygmomanometer.* Proc. of Intern. Conf. on Appl. of Inf. and Com. Tech. and Statist. in Ec. and Educ., October – 2012, UNWE, Sofia.
7. Э. Сейдж, Дж. Мелс. *Теория оценивания и ее применение в связи и управлении.* Москва, "Связь", 1976, 496 с.

Implementation of Web-Based System for Short-Term Forecasting of Energy Consumption

Silyan Arsov[1], **Kiril Sirakov**[1] and **Denislav Arsov**[2]

[1]University of Russe, Studentska str. 8, 7017 Ruse, Bulgaria
si.arsov@gmail.com, csirakov@uni-ruse.bg
[2]Walmart, 1400 Lead Hill Blvd, Roseville, CA 95661, USA
denis_arsov@yahoo.com

Abstract. The necessity for development of automated web-based software system for short-term forecasting of electricity consumption in the separate production companies has been justified. The author's generalized algorithms of programming procedures for calculating the forecasted value of the measured variable – electricity, by the methods of least squares and exponential average are examined in the research. The experimental results for a period of ten days, obtained by the work of the software system, which was developed using the created algorithms are presented. Real data of measured electricity consumption in a regular operating manufacturing company are used.

Key words: Electrical energy, Forecasting, Planning, Web based system.

1 Introduction

Electricity is the most convenient way for transportation and conversion into other types of energy among the existing energy carriers. The time of electricity production practically coincides with the time of consumption. Unlike almost all other energies it cannot be stored, collected or accumulated in significant quantities. Because of these specific features of electricity it is necessary to balance not just the total electrical energy quantity but also the delivered and consumed power at any point of the time. This is accomplished through precise planning of its production, distribution and consumption.

At each stage of the electricity path from producers to consumers various economic subjects are situated. This requires accurate forecasting of consumption in the separate elements of the energy grid for the corresponding periods of time.

The only way to make an objective forecast is by using methods of mathematical statistics.

The mathematical statistics methods require a significant amount of calculations. The lack of automation of the computing process and due to the execution of a large number of mathematical operations are prerequisites to different types of errors occur.

To achieve effective practical application of the statistics methods it is necessary to develop automated hardware and software for mathematical processing of statistical data. In the proposed research we examine web-based software implementation of an approach which develops short-term forecasts for planning purposes of the electricity consumption.

There are many universal software systems to solve statistics problems of this type [3], [4], but unlike them, the proposed web-based system enables the clients to access it through the Internet, regardless of the location of their work place and without having to install the software system to their work stations.

To achieve the objective of the project, the following tasks are resolved: the main components of the architecture of the software system are defined [1]; generalized algorithms for the execution of the calculation process, computing the forecasted data by the method of least squares (MLS) and method of exponential average (MEA) are designed; programming procedures for the execution of the created algorithms are designed and implemented, they make it possible to calculate the forecasted data, based on actual reported information of electricity consumption for a prior period; the structures of the data elements involved in the calculation process are defined; programming procedures, which provide options to enter, modify, delete and store the data in files, output information to the screen and print it, are designed and implemented; the necessary numbers of forms are developed to provide the user interface with the processes of input, output, editing, deleting and storing of used data.

2 Generalized algorithm of the computing procedure to obtain the forecasted data of the measured variable - electricity using the method of least squares (MLS)

To calculate the forecasted values of the observed variable, electricity consumption, is achieved by a procedure, implemented by selected programming language, which execute various elements of the method of least squares [1], [2], [5] as compiled and presented below summary algorithm.

2.1 Input parameters of the procedure

- Base period $i = 1 \ldots N$ (in this case $i = 1 \ldots 20$);
- Measured values during the base period ($i = 1 \ldots 20$);
- Forecast period $j = 1 \ldots L$ (in this case $j = 1 \ldots 10$);
- Margin of error α (risk of error), adequate enough forecast is obtained with $\alpha = 0,1$ in most cases.

2.2 Output parameters of the procedure

- Calculated values of the straight line in the graphs - $\hat{Y}_i$;
- Upper critical limit of the forecast;
- Forecasted value for the corresponding period;
- Lower critical limit of the forecast.

2.3 Sequence of the computing operations

1: Calculation of the parameters a_0 and a_1 in the polynomial equation:

$$\hat{Y}_i = a_0 + a_1 . t_i \ , \qquad (1)$$

in the following terms:

$$a_1 = \frac{N.\sum_{i=1}^{N} Y_i.t_i - \sum_{i=1}^{N} Y_i . \sum_{i=1}^{N} t_i}{N.\sum_{i=1}^{N} t_i^2 - (\sum_{i=1}^{N} t_i)^2} \qquad a_0 = \frac{\sum_{i=1}^{N} Y_i}{N} - a_1 \frac{\sum_{i=1}^{N} t_i}{N} \ , \qquad (2)$$

where t_i is the corresponding base period, ($i = 1...N$).

2: Graphs construction for the base period of Y_i and $\hat{Y}_i$,

where:

- Y_i is actual values reported in the base period;
- $\hat{Y}_i$ - values for the base period, mathematically derived from (1).

3: Calculation of the deviation:

$$\varepsilon_i = Y_i - \hat{Y}_i \ . \qquad (3)$$

4: Equation for calculation of the standard error:

$$S_Y = \sqrt{\frac{\sum_{i=1}^{N} \varepsilon_i^2}{N-1}} \ . \qquad (4)$$

5: Calculation of the average forecasted value:

$$\hat{Y}_{N+j} = a_0 + a_1 . t_{N+j} \ , \qquad (5)$$

where t_{N+j} is the corresponding forecasted period, ($j = 1 \ ... \ L$).

6: The L number of values for the variable $K^*_{N,j,\alpha}$ are read from the table previously stored in a file.

7: Calculation of the critical limits of the forecast confidence interval:

$$\hat{Y}_{N+j} - K^*_{(N,j,\alpha)}.S_Y \le \hat{Y}_{N+j} \le \hat{Y}_{N+j} + K^*_{(N,j,\alpha)}.S_Y \ , \qquad (6)$$

where:

- $\hat{Y}_{N+j} + K^*_{(N,j,\alpha)}.S_Y$ is the upper critical limit of the forecast;
- $\hat{Y}_{N+j} = a_0 + a_1 . t_{N+j}$ - estimate value for the corresponding period;
- $\hat{Y}_{N+j} - K^*_{(N,j,\alpha)}.S_Y$ - the lower critical limit of the forecast.

8: Representation in a graphic form of the forecasted values and confidence intervals obtained by calculations.

3 Results obtained using the experimental software system for forecasting energy consumption of the manufacturing company for the last ten days of the month, by applying MLS

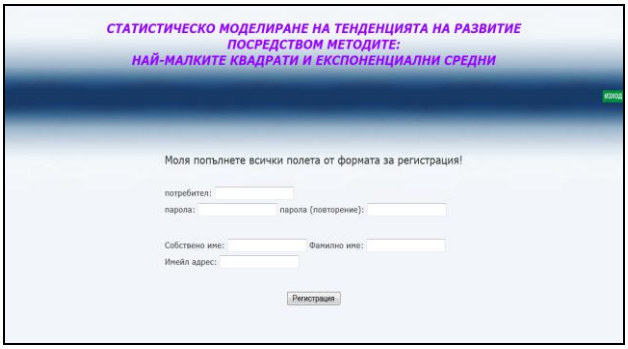


Fig. 1. Application Home Page.

3.1 Table of measured and calculated values using MLS for energy consumption (Table 1)

Table 1. Measured and calculated values for energy consumption by manufacturing company using MLS.

MEASURED AND CALCULATED VALUES USING MLS					
Base period, [days]	Measured values, [kWh]	Calculated values, [kWh]	Base period [days]	Measured values, [kWh]	Calculated values, , [kWh]
1	19 368	32 194	11	167 644	168 791
2	39 966	45 854	12	187 842	182 450
3	60 174	59 513	13	189 764	196 110
4	80 214	73 173	14	203 220	209 770
5	97 950	86 833	15	217 759	223 429
6	103 018	100 492	16	233 285	237 089
7	115 908	114 152	17	252 493	250 749
8	129 992	127 812	18	272 065	264 408
9	143 398	141 471	19	282 893	278 068
10	157 388	155 131	20	284 875	291 728

3.2 Table of calculated forecast values applying MLS for energy consumption (Table 2)

Table 2. Forecasted values, calculated applying MLS.

FORECAST VALUES, CALCULATED USING **MLS**			
Forecast period, [days]	Forecast value, [kWh]	Lower limit, [kWh]	Upper limit, [kWh]
21	305387	293885	316890
22	319047	307389	330705
23	332707	320881	344533
24	346366	334362	358371
25	360026	347832	372220
26	373686	361292	386080
27	387345	374742	399949
28	401005	388183	413827
29	414665	401615	427715
30	428324	415039	441610

3.3 Graphical representation of the calculation of the forecasted values using MLS

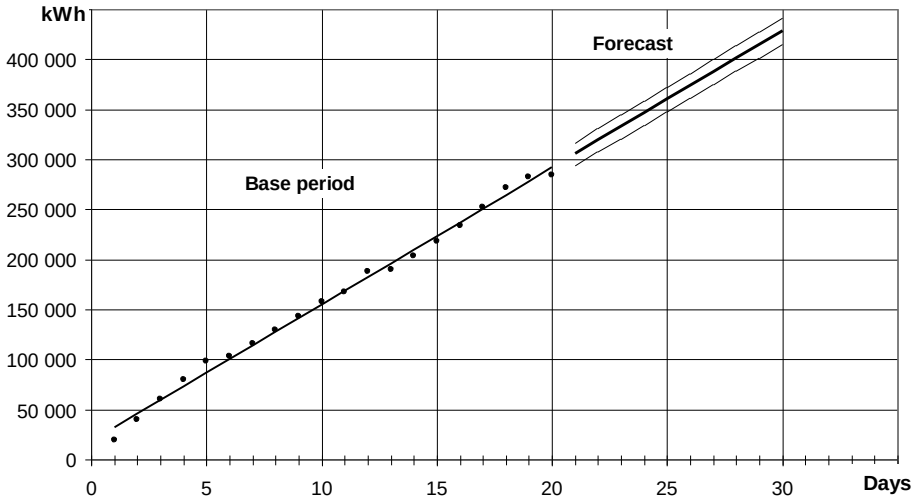


Fig. 2. Graphical representation of the forecasted values calculations using MLS.

4 Summarized algorithm of the procedure for calculating the forecast values of the measured variable – electricity, using the method of exponential average (MEA)

To calculate the forecasted values of the observed variable- electricity consumption, a procedure is implemented on a selected programming language. The procedure executes the different mathematical expressions of the exponential average method [1], [2], [5] according to the following summarized algorithm:

4.1. Input parameters of the procedure

- Base period ***i = 1 .. N*** (in this case ***i = 1 .. 20***);
- Measured values during the base period (***i = 1 ... 20***);
- Forecast period ***j = 1 ... L*** (in this case ***j = 1 ... 10***);
- Margin for error α, in most cases **α = 0,1** (admissible risk of error to 10%);
- Coefficient ***d***, which determines the influence of the set members. When its value is close to 0 then calculating the forecast, the influence of all set members of the base period is accounted. When the value of ***d*** is close to 1 then the last set members influence mainly the forecasting process. Most often ***d*** is chosen in the range (0,2 ... 0,5).

4.2 Output parameters of the procedure

- Calculated values of the straight line in the graphs - $\hat{Y}_i$;

- Upper critical limit of the forecast;
- Forecasted value for the corresponding period;
- Lower critical limit of the forecast.

4.3. Sequence of the calculation operations

1: Calculation of the parameters a_0 and a_1 using the first few periods, usually the first three (i.e., Y_1, Y_2, Y_3):

$$a_1 = \frac{N.\sum_{i=1}^{N} Y_i.t_i - \sum_{i=1}^{N} Y_i . \sum_{i=1}^{N} t_i}{N.\sum_{i=1}^{N} t_i^2 - (\sum_{i=1}^{N} t_i)^2} \; ; \qquad a_0 = \frac{\sum_{i=1}^{N} Y_i}{N} - a_1 \frac{\sum_{i=1}^{N} t_i}{N} \; , \tag{7}$$

where:

- N is the base period;
- Y_i - measured values in the base period;
- t_i - the corresponding period.

2: Calculation of the initial exponential averages S_0^1 and S_0^2 using the expressions:

$$S_0^1 = a_0 - \frac{1-d}{d}.a_1 \; ; \qquad S_0^2 = a_0 - \frac{2.(1-d)}{d}.a_1 \; . \tag{8}$$

3: Calculation of each of the values $S_i^1, S_i^2, a_0, a_1, \hat{Y}_i$ and the variance ε_i according to mathematical expressions in (9), (10), (11) and (12) for each $i = 1 \ldots N$, (for the corresponding case $i = 1 \ldots 20$):

$$S_i^1 = d.Y_i + (1-d).S_{i-1}^1 \; ; \qquad S_i^2 = d.S_i^1 + (1-d).S_{i-1}^2 \; , \tag{9}$$

$$a_{0,i} = 2.S_i^1 - S_i^2 \; ; \qquad a_{1,i} = \frac{d}{(1-d)}.(S_i^1 - S_i^2) \; , \tag{10}$$

$$\hat{Y}_i = a_{0,i} + T.a_{1,i} \; ; \qquad T=1 \tag{11}$$

$$\varepsilon_i = Y_i - \hat{Y}_i \; . \tag{12}$$

4: Construction of two graphs: the base period with the read actual values of the measured variable and the forecasted period with the values obtained from the calculation of expression (11) - $\hat{Y}_i$

5: Standard error calculation:

$$S_Y = \sqrt{\frac{\sum_{i=1}^{N} \varepsilon_i^2}{N-1}} \quad . \qquad (13)$$

6: ***L*** times - calculation of the forecast error (***j = 1 ... L***, for the present case - 10 times):

$$\sigma_{\hat{Y},N+j} = S_Y . \sqrt{\frac{d}{(2-d)^3} . (10 - 14.d + 5d^2 + 2.d.j.(4 - 3.d + d.j))} \quad . \qquad (14)$$

7: From the table, which is stored in a file in advance, the value of the Student's ***t*** distribution $t_{(N-2),\alpha}$ is read, with degrees of freedom ***N-2*** and margin of error α.

8: Calculation of the average forecasted value:

$$\hat{Y}_{N+j} = a_{0,N} + a_{1,N} . t_j \quad , \qquad (15)$$

where t_j are the values of the forecasting period (for the specific area t_j ***= 1 .. 10***).

9: Calculation of the confidence interval of the forecast (for the specific subject area is necessary to get 3 rows with 10 values):

$$\hat{Y}_{N+j} - \sigma_{\hat{Y},N+j} . t_{(N-2),\alpha} \le \hat{Y}_{N+j} \le \hat{Y}_{N+j} + \sigma_{\hat{Y},N+j} . t_{(N-2),\alpha} \quad , \qquad (16)$$

where:

- $\hat{Y}_{N+j} + \sigma_{\hat{Y},N+j} . t_{(N-2),\alpha}$ is the upper critical limit of the forecast;
- $\hat{Y}_{N+j} = a_{0,N} + a_{1,N} . t_j$ - forecasted value for the corresponding period;
- $\hat{Y}_{N+j} - \sigma_{\hat{Y},N+j} . t_{(N-2),\alpha}$ - the lower critical limit of the forecast.

10: Representation in graphic form of the forecast values and confidence intervals obtained by calculations.

5 Results obtained by forecasting the energy consumption of the manufacturing company for the last ten days of the month by applying the MEC

5.1 Table of measured and calculated values applying MEA (Table 3)

Table 3. Measured and calculated values by applying MEA.

MEASURED AND CALCULATED VALUES USING **MEA**					
Base period, [days]	Measured values, [kWh]	Calculated values, [kWh]	Base period, [days]	Measured values, [kWh]	Calculated values, [kWh]
1	19 368	19433	11	167 644	175429
2	39 966	39797	12	187 842	186008
3	60 174	60296	13	189 764	201658
4	80 214	80635	14	203 220	209236
5	97 950	100784	15	217 759	219270
6	103 018	119447	16	233 285	231466
7	115 908	129698	17	252 493	245524
8	129 992	140054	18	272 065	262835
9	143 398	151406	19	282 893	282131
10	157 388	163084	20	284 875	297176

5.2 Table of calculated **forecast** values applying MEA for energy consumption (**Table 4)**

Table 4. Forecased values,calculated applying MEA.

FORECASTED VALUES, CALCULATED USING **MEA**			
Forecast period, [days]	Forecast value, [kWh]	Lower limit, [kWh]	Upper limit, [kWh]
21	304452	294652	314253
22	318002	306873	329131
23	331551	319068	344035
24	345101	331245	358957
25	358650	343409	373892
26	372200	355563	388837
27	385749	367709	403790
28	399299	379849	418749
29	412849	391985	433712
30	426398	404117	448680

3.3 Graphical representation of the calculation of the forecasted values applying MEA

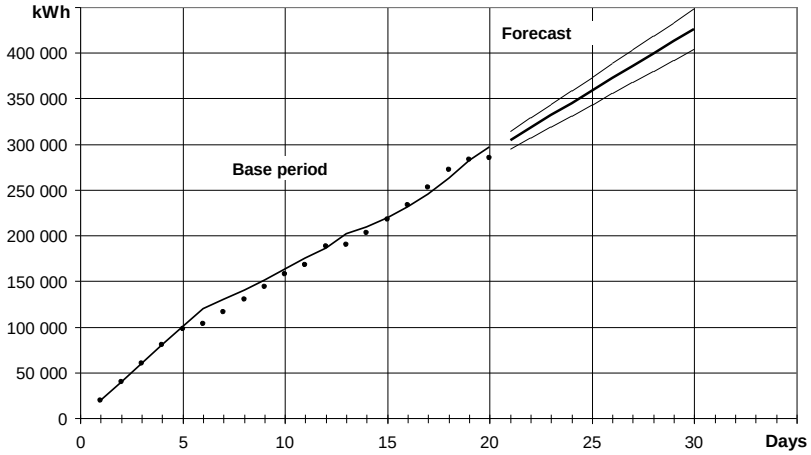


Fig. 3. Graphical representation of the calculation of the forecasted values applying MEA.

6 Conclusion

Algorithms for the realization of forecasts methods of least squares and exponential averages are presented.

A project and implementation of a web based software system for mathematical processing of statistical data and subsequent generation of forecast data in tabular and graphical form are proposed.

The web based automated software system was developed using the PHP programming language and data bases management system MySQL.

It was tested with data on consumption of electricity produced by Bulgarian industrial enterprises. Some of these companies are participants in the liberalized energy market in the country.

References

1. Arsov, S., K. Sirakov: An Approach for Development of Web Based System for Electrical Energy Consumption Planing. In: Proceedings of the International Conference on Application of Information and Communication Technology in Economy and Education, (ICAICTEE-2011), UNWE, Sofia, Bulgaria, pp. 264-272 (2011)
2. Brase, Ch., Brase, C.: Understandable Statistics: Concepts and Methods. Brooks/Cole Publishing Company, CA, USA (2009)
3. Krastev, G., Georgiev, Ts.: Tools for Automaton of the Science Researches. Avangard print, Ruse (2002)
4. Mitkov, A.: Theory of the Experiment. Dunav Press, Ruse (2011)
5. Velichkova, N.: Statistical methods for researching and forecasting the development of the social- economical events. Nauka i Izkustvo, Sofia (1981)
6. Welling, L., Thomson, L.: PHP and MySQL Web Development. Pearson Education (2003)

Lattice-based cryptographic algorithms and their applications in protection of business-related data processing

Angelin Lalev[1]

[1] D. A. Tsenov Academy of Economics - Svishtov, Bulgaria
lalev@uni-svishtov.bg

Abstract. Lattice-based cryptography is a subject of intensive research for about a decade and a half. While the research is, and probably will be ongoing for quite a long time, the first results in the form of cryptographic algorithms with unique properties are already available. One of the interesting properties of some lattice-based cryptographic schemes is that they are not known to be broken by quantum algorithms. Another group of lattice-based cryptographic schemes implements the so called "fully homomorphic encryption", which allows processing directly on encrypted data and holds the promise eventually to solve the security problems with cloud computing. Unfortunately, the recent results in the area are relatively unknown in Bulgaria both in practice and in the academia. We believe that this necessitates more wide exposure and popularization of the underlying ideas. This would ultimately help specialists in business data processing by assisting them in monitoring and assessing eventual future developments and breakthroughs in these important technologies, which without any doubt will bring profound changes in the way business data is processed.

Keywords: point lattice, lattice-based cryptography, GGH, homomorphic encryption, privacy homomorphisms, NTRUSign, NTRUEncrypt, quantum computing, security

Motivation behind research in lattice-based cryptography

Lattice-based cryptography is a relatively new approach towards cryptographic protection of data in computer systems. It is a counterpart of more widely known, thoroughly tested and smoothly-working traditional algorithms (such as RSA, DSA, AES, IDEA), which seem so far to fulfill their purpose more than adequately. Some lattice-based cryptography schemes though, have exhibited unique properties which allow their application in completely new scenarios, very relevant to the current developments in computer technology.

A very desirable property of lattice based cryptography is related to the fact that there are lattice-based schemes, which are resilient to cryptanalysis, conducted with the help of quantum computers. Although quantum computers are still at early research phase, they are defined quite rigorously as machines that use the phenomena of quantum superposition to process information, stored in the so called "quantum bits" or "qubits". By the nature of quantum mechanics such qubits (which can be realized in theory by using different elementary particles and their characteristics), can be in both basic states, corresponding to 0 and 1 at the same time. This allows creation of algorithms for quantum computers, which defy the complexity hierarchy established for the classical Turing machine model.

This fact is especially important for the field of cryptography due to work by Shor[1], who created an algorithm for factoring numbers into prime factors, which will work in polynomial time[18] on quantum computer. The most important cryptographic algorithms, used today base their security on the supposed intractability of the exact same problem[19]. Eventual breakthrough in quantum computer technology would then mean the loss of the most popular today cryptographic schemes like DSA, RSA, ECIES, ECDSA. Without these schemes the fundamental building blocks of the modern e-business, e-commerce and e-government, namely secure payments, secure browsing of web, secure e-mail, digital signatures, secure access to servers and web-services will be seriously hampered.

Such breakthrough in quantum computer technology then will require quick change of encryption algorithms, which given the scales of their adoption, cannot happen in an instant[20]. In this respect, the research and implementation of lattice based cryptographic schemes is an insurance against potential and very sudden changes in our current understanding of the computational complexity of the factoring problem. The most studied lattice schemes, which support asymmetric cryptography and digital signatures are NTRUEncrypt and NTRUSign, one of which will be overviewed in this paper.

Lattice-based cryptography has another very interesting application. It has been demonstrated that lattice-based schemes may be modified to achieve the so called "fully homomorphic encryption". Fully homomorphic encryption allows processing on encrypted data. Traditional encryption algorithms (like the ones based on Feistel networks), rely on scrambling the data so it is indistinguishable from a random sequence of numbers to an observer, who doesn't have the decryption key. So it is impossible for such observer to perform any meaningful data processing operations which affect the contents of the encrypted data. Yet such property is extremely desirable from the standpoint of cloud computing.

Cloud computing is one of the very successful ideas in the last several years, which combines virtualization with outsourcing to allow the organizations to outsource their data processing operations to specialized data centers, where serious economies of scale can be realized due virtualization capabilities of the modern computer hardware. Cloud computing has only one flaw, which is to some extent subjective, but nonetheless – very serious. It concerns security of the information. Since cloud does involve outsourcing of the data, the data often crosses the limits not only of the organization, but national borders and the borders of political unions like EU. By crossing them, it not only falls under jurisdiction of other countries but it is exposed to new threats to its security, which concern the good will and competence of the cloud provider. (See [5],[6] for some real-world cases, concerning

[18] In short, polynomial time is a term that represents a class of algorithms, whose worst-case running time is polynomial function of the size of the input. This is contrasted to the non-polynomial time algorithms, which are considered too computationally intensive to be efficiently implemented on traditional computers, conforming to the Turing model, except for very small size of the input.

[19] The best known algorithm for factoring numbers on classical computers – the general number field sieve[2] – runs in subexponential time, which makes breaking the modern encryption schemes intractable task on even most powerful supercomputers.

[20] The story of the still most popular cryptographic hash function – the MD5 algorithm – suggest probable scenario. Although very serious weaknesses in MD5 were discovered in the previous decade, rendering it very insecure, its ubiquitousness still keeps it in use even when very serious computer viruses are starting to exploit its weaknesses to forge digital signatures (See [3],[4] for amazing story).

outsourcing as a whole). This is often undesirable, which presently limits the universal adoption of cloud computing. Fully homomorphic encryption would in theory allow the cloud providers to process encrypted data, whose true meaning remains fully inaccessible to them. This in turn would lead to much more massive adoption of cloud computing technologies in the practice of all organizations. In 2009, thanks to the work of Gentry [7],[8] the first ever fully homomorphic scheme has been proposed. It is based on lattices and although it introduces almost unbearable computational overhead, there are great hopes that lattices, due their deep combinatorial structure, would provide more adequate solutions to the problem in time.

Mathematical background on lattices

Point lattices[21] are algebraic structures, which can be described as discrete additive subgroups of some n-dimensional Euclidean space R^n. Alternatively, more convenient equivalent definition describes them as the set of all linear combinations with integral coefficients of some set of linearly independent vectors in R^n. This set is called basis of the lattice and is conveniently represented as matrix, whose columns are the components of the vectors in the base. In mathematical notation, the base B may be written as

$B = \{\boldsymbol{b_1}, \boldsymbol{b_2}, \dots, \boldsymbol{b_m}\}, m \leq n$

where $\boldsymbol{b_1}, \boldsymbol{b_2}, \dots, \boldsymbol{b_m}$ are the vectors of the base and n is the dimension of the vector space[22]. The lattice, formed by all linear combinations of the vectors in B may be denoted as

$$\mathcal{L}(B) = \left\{ \sum_{i=1}^{m} c_i \boldsymbol{b_i} : c_i \in Z \right\}$$

In matrix notation $B = [\boldsymbol{b_1} \vdots \boldsymbol{b_2} \vdots \cdots \vdots \boldsymbol{b_m}]$ and $\mathcal{L}(B) = \{B\boldsymbol{c} : \boldsymbol{c} \in Z^m\}$.

Each lattice has infinitely many bases. For each two of them B_m and B_n there exists unimodular matrix such that $B_n \boldsymbol{U} = B_m$. (See fig.1)

[21] There is unfortunate confusion of names concerning point lattices. In practically all relevant research, and especially in the field of computer science, they are known just as "lattices". This name conflicts with a term in order theory. Fortunately, order theory has not direct applications to cryptography, so the use of the term is clear from the context. We will conform to the established term and call point lattices just "lattices" from here on.

[22] Cryptography normally uses only full rank lattices $(n = m)$, whose basis vectors contain only integral components.

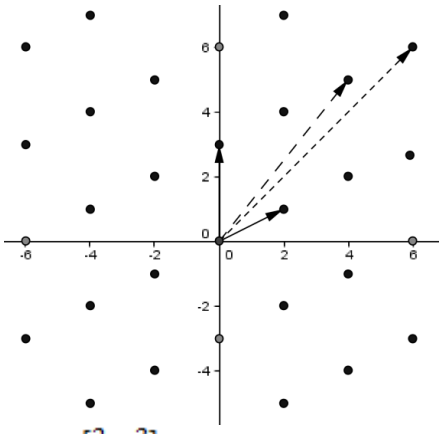


Fig. 1. Lattice in R^2 with base $\begin{bmatrix} 2 & 3 \\ 1 & 0 \end{bmatrix}$. Alternatively the same lattice can be represented by base $\begin{bmatrix} 4 & 6 \\ 5 & 6 \end{bmatrix}$ (the vectors with dashed line) and even base $\begin{bmatrix} 246088062 & 12304103 \\ 41001 & 2050 \end{bmatrix}$.

The ideal lattices are special kind of lattices which play serious role in homomorphic encryption schemes. Ideal lattices consist of vectors, that correspond to some ideal in quotient ring of type $Z[x]/f(x)$, where $Z[x]$ is the ring of all polynomials with coefficients in Z and $f(x)$ is a monic polynomial of degree n with integer coefficients (usually taken to be $x^n - 1$ or $x^n + 1$). The members of this quotient ring are polynomials of degree at most $n - 1$ which may be identified with vectors in n-dimensional Euclidean space. If the ideal is principal (generated by one element of the ring), there is very easy way of generating good base for the corresponding lattice. For each ideal $I = (v)$, $v \in Z[x]/f(x)$, the so called "rotational base" may be calculated by the formula

$v_i = v \times x^{i-1} \bmod f(x)$ for $i \in \{1, n\}$.

The lattices are closed with respect to the usual binary operations of addition and negation of vectors and ideal lattices are closed under multiplication modulo $f(x)$.

Two important algorithmic problems, related to lattices are the problems for finding the shortest vector (abbreviated as "SVP") and the problem of finding the closest vector (abbreviated as "CVP"). SVP concerns finding the shortest vector of lattice, given base of the lattice which has long and nearly collinear vectors. The problem in multiple dimensions[23] is a NP-hard problem [9],[10]. This means that with increasing the number of dimensions of the lattice and choosing bad enough base for the lattice it becomes incredibly hard to find shortest vector in the lattice from computational point of view. This property is used as the major building block of lattice-based cryptography. CVP is closely related to SVP problem. It gives a vector and a lattice such that the vector is not in the lattice. It then asks for the vector in the lattice which is closest to the original point. It is shown that an algorithm that solves CVP can be modified to solve SVP as easily from computational point of view [11].

Since both problems seem to be intractable in higher dimensions, the best alternative of finding the exact short vector is to use approximation. The best known algorithm that gives

[23] The cases where the lattice and the vectors in the base are orthogonal, are of course, excluded, since they are the easiest special case from algorithmic point of view and calculation of SVP and CVP with them are trivial.

such approximation is the celebrated LLL algorithm [12]. It can only give relatively short vector, which is within $\left(\frac{2}{\sqrt{3}}\right)^n$ of the length of the shortest vector (where n is the rank of the lattice). Lattice cryptography entirely depends on the hardness of finding short vectors in lattices, so LLL is a major tool for cryptanalysis. Fortunately, LLL can be defeated with choosing carefully bad bases and increasing the number of dimensions in the range of 1000.

Geometrically, the basis of a full-rank lattice defines a half open n-dimensional parallelotope (we will call it base parallelotope) in R^n

$$\mathcal{P}(B) = \left\{x_i \boldsymbol{b_i} : x_i \in \left[-\frac{1}{2}, \frac{1}{2}\right)\right\}$$

Each lattice with integral vectors of the base gives equivalence relation E on Z^n, which may be defined as $(\boldsymbol{x},\boldsymbol{y}) \in E$ (for two vectors $\boldsymbol{x} \in Z^n, \boldsymbol{y} \in Z^n$) if and only if $(\boldsymbol{x} - \boldsymbol{y}) \in \mathcal{L}(B)$. This relation splits Z^n into different equivalence classes. Each of these classes has exactly one unique representative in $\mathcal{P}(B)$.

This unique representative may be calculated by the formula[24]

$$\boldsymbol{y} = \boldsymbol{x} - B\lfloor B^{-1}\boldsymbol{x} \rceil$$

Finding such unique representative is called reduction of x modulo $\mathcal{L}(B)$ (denoted $x \bmod \mathcal{L}(B)$). The base parallelotope $\mathcal{P}(B)$ circumscribes a n-dimensional ball $O(B)$. Generally, the closer the vectors of the chosen base are to orthogonality, the bigger the radius of $O(B)$ is.

Lattice-based homomorphic encryption

The scheme proposed by Genrty aims to achieve addition modulo 2 of two encrypted bits and multiplication modulo 2 of two encrypted bits. These operations are equivalent to OR and AND gates, which in turn may be used in combinations to compute any computable function (program) much like this is done in electronic circuits. To achieve this, the scheme uses ideal lattices. The main idea behind the scheme is to generate a "good" and a "bad" base of an ideal lattice. The bad base will have considerably longer (and therefore - nearly collinear vectors) than the good base and therefore the inscribed ball in its base parallelotope will be considerably smaller than its counterpart in the good base. The good base will play the role of the private key and will be used for decryption of the data and the bad base would be useful only for encryption of the data and thus can be distributed as public key.

The good base in Gentry's scheme is generated by choosing a polynomial from $Z[x]/(x^n + 1)$ and generating rotational base B_g whose $O(B_g)$ has some radius r_g. The bad base B_b will be then calculated as Hermite normal form of the good basis's matrix. With proper choice of parameters, this will guarantee, that the basis parallelotope will be strongly "skewed" and the radius r_b of $O(B_b)$ will be much smaller then r_g. These settings allow creation of one-way cryptographic function[25]. Such cryptographic function includes a padding scheme, which maps each of the two possible values of the bit, being encrypted, to

[24] $\lfloor \ \rceil$ denotes Gaussian rounding.

[25] One that is easy to calculate, but has inverse that is very hard to calculate.

one of two collinear vectors (we denote them by m_0 for the vector that corresponds to 0 and m_1 for the vector that corresponds to 1), whose length is taken to be considerably smaller then r_g and bigger than r_b. Adding one of these vectors (corresponding to the bit to be encrypted), to a random vector in the lattice will result a vector ψ outside the lattice. For completion of this operation only the public key B_b is needed (B_b is multiplied by random integer vector and the vector, corresponding to 0 and 1 is added to the product). To recover the 0 or 1 from ψ, one must find the closest vector to ψ in the lattice and subtract it from ψ. Finding the closest vector though is hard computationally, so the described operation, is in fact encryption. To decrypt properly ψ it is only necessary to do reduction modulo $\mathcal{L}(B_g)$. Due to the padding scheme, we have guarantees, that ψ will be at no more than r_g from a vector in the lattice, which gives assurance that the resulting vector will be m_0 or m_1, depending on the value of the bit, encrypted in ψ. The encryption has other interesting property. If ψ_1 encrypts some bit $\pi_1$and ψ_2 encrypts some bit π_2, then $\psi_1 + \psi_2$ encrypts $\pi_1 + \pi_2$, given it's length does not surpass r_g. This allows encrypted addition of bits. The scheme has the same properties towards multiplication of bits, which easily gives the desired AND and OR operations on encrypted data[26]. The problem with this approach, and where the greatest contribution of Gentry lays, is that after several operations on encrypted data, the length of the resulting vector ψ will surpass r_g and reduction modulo $\mathcal{L}(B_g)$ will give incorrect decryption. To achieve improvement, Gentry proposes approach to modify parameters of the scheme, so the decryption (calculation of reduction modulo $\mathcal{L}(B_g)$) will be simple enough to be presented as a sequence of AND and OR gates, such that executing them in on encrypted bits won't make ψ to surpass r_g. He called this "squashing" of the decryption circuit. Gentry then observed that any scheme that can be squashed can be turned into fully homomorphic encryption, by making the side, that computes on encrypted data to compute the operations for decryption on the data after application of only one logical gate (AND, OR). This can be done with encrypted version of B_g (so the side that does calculations on encrypted data never gets the unencrypted version of B_g). The result will be new ciphertext ψ that is shorter and encrypts the same data. He called this "bootstrapping". By reducing the potential error, Gentry then proved, that such scheme can do unlimited number of AND and OR gates at the cost of greatly increasing the overhead.

Lattice-based asymmetric encryption

NTRUEncrypt [14] is a proposed family of asymmetric cryptosystems that concern operations in polynomial ring $Z[x]/(x^n - 1)$ with the usual multiplication, which is described by the cyclic convolution product.

[26] Describe so far is a slight modification of the so called GGH scheme [13], which is used by Gentry. The approach by Gentry though is more important for the bootstrapping and squashing processes, which are too long to be described in such a short exposition.

$$h_k = \sum_{i+j \equiv k_{(mod\ n)}} f_i g_j$$

Where **h**, **f**, and **g** are polynomials in $Z[x]/(x^n - 1)$, which may be identified with vectors. Each of the cryptosystems in the family may be characterized by three numbers (n, p, q) and four sets $(\mathcal{L}_f, \mathcal{L}_g, \mathcal{L}_\pi, \mathcal{L}_r)$ of polynomials. n is prime and p and q are coprime, q being much more larger than p. The public key in NTRUEncrypt is generated by choosing two polynomials $\boldsymbol{F}$ and $\boldsymbol{g}$ at random. $\boldsymbol{F}$ and $\boldsymbol{g}$ must be of degree at most $n-1$ and with coefficients of each of the terms -1, 0 or 1. From **F**, another polynomial $\boldsymbol{f}$ must be chosen such that it satisfies two conditions. First, $\boldsymbol{f} = 1 + p\boldsymbol{F}$ (which implies that inverse of $\boldsymbol{f}$ modulo p is 1) and second - that $\boldsymbol{f}$ is invertible modulo q. The public key $\boldsymbol{h}$ is then generated by the formula $\boldsymbol{h} \equiv p * \boldsymbol{f_q} * \boldsymbol{g}_{(mod\ q)}$ where $\boldsymbol{f_q}$ is the inverse of $\boldsymbol{f}$ modulo . $\boldsymbol{f}$ is then taken as private key.

To encrypt message for side B, side A must know it's public key $\boldsymbol{h}$. It then maps the plaintext by some padding scheme to polynomial $\boldsymbol{\pi}$, having coefficients -1,0 or 1. A then generates random polynomial $\boldsymbol{r}$ with relatively small coefficients (not necessarily limited to -1,0 or 1, but much more smaller than q) and calculates the ciphertext $\boldsymbol{\psi} \equiv \boldsymbol{r} * \boldsymbol{h} + \boldsymbol{\pi}_{(mod\ q)}$. (**r** is kept secret by side A and is not transmitted with the ciphertext. If not, then anybody can calculate $\boldsymbol{r} * \boldsymbol{h}_{(mod\ q)}$ and decrypt easily $\boldsymbol{\pi}$).

The holder of the private key can then decrypt the message by first calculating

$$\boldsymbol{a} \equiv \boldsymbol{f} * \boldsymbol{\psi}_{(mod\ q)} \equiv \boldsymbol{f} * \boldsymbol{r} * \boldsymbol{h} + \boldsymbol{f} * \boldsymbol{\pi}_{(mod\ q)} \equiv$$

$$\equiv \boldsymbol{f} * \boldsymbol{r} * p * \boldsymbol{f_q} * \boldsymbol{g} + \boldsymbol{f} * \boldsymbol{\pi}_{(mod\ q)} \equiv \boldsymbol{r} * p * \boldsymbol{g} + \boldsymbol{f} * \boldsymbol{\pi}_{(mod\ q)}$$

When calculating **a**, the values for coefficients of **a** are taken from the interval $[-q/2, q/2]$ instead from $[0, q-1]$ to allow proper decoding of the message (i.e. coefficient from $(\frac{q}{2}, q-1]$ is represented as inverse modulo q of smaller coefficient from [1 to q/2)). The choices of $\boldsymbol{f}, \boldsymbol{g}, p$ an $\boldsymbol{r}$ and q are chosen so (q is much bigger than coefficients of the calculated polynomials), that reduction modulo q of a won't change the coefficients of $\boldsymbol{a}$. Then reduction of $\boldsymbol{a}$ modulo p will yield $\boldsymbol{\pi}$.

NTRUEncrypt was thought as encryption system on polynomial rings, but today is considered lattice problem, because it turned out that inverting of the underlying one-way function is equivalent of finding the shortest vector in a special lattice [15]. NTRUEncrypt is easily converted into signature scheme, which facilitates it's usage for digital signatures too.

Conclusion

Lattice-based encryption solves problems that are strategic and will have more impact in the long-term. From practical standpoint this is the domain of the specialists of business information systems, who are concerned with the strategic aspects of the information processing of the organization. The lattice-based cryptography relies on much more complex and new results than traditional cryptography, which makes the recent results hardly accessible to such specialists. This necessitates help from academia which in Bulgaria is not available yet. In practice there are no publications in Bulgarian exist on the topic, nor are the

practitioners informed about the new status of NTRU-based encryption as a ASC X9 standard (namely X9.98). Correcting such situation should not be very hard though. The standard curriculum of the specialists on business-oriented data processing already includes enough material on mathematics (due to operations research oriented mathematical education) and enough material on applied cryptography to be easily modified.

References

1. SHOR, P. "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer", Society of Industrial and Applied Mathematics Journal of Computing (SICOMP) 26**,** pp. 1484-1509 (1997).
2. POMERANCE, C. "A Tale of Two Sieves", Notices of the AMS 43 (12): pp. 1473–1485 (1996)
3. Crypto breakthrough shows Flame was designed by world-class scientists, ArsTechnica Blog, http://arstechnica.com/security/2012//06/flame-crypto-breakthrough/ (Jun 2012)
4. Researchers find direct link between Flame, Stuxnet malware, ComputerWorld (Jul 2012)http://www.computerworld.com/s/article//9227954/Researchers_find_direct_link_between_Flame_Stuxnet_malware
5. Microsoft admits Patriot Act can access EU-based cloud data, ZDNET Blog, http://www.zdnet.com/blog/igeneration/microsoft-admits-patriot-act-can-access-eu-based-cloud-data/11225
6. "Massive" Credit Card Breach of Estimated 10 Million Accounts: Where Are Those Smart Cards?, Forbes, http://www.forbes.com/sites/anthonykosner/2012/03/31/massive-credit-card-breach-of-estimated-10-million-accounts-where-are-those-smart-cards/
7. GENTRY, C. Fully homomorphic encryption using ideal lattices, In proc. of 41-st Symposium on the Theory of Computing (STOC), 2009, pp. 169-178.
8. GENTRY, C. A fully homomorphic encryption scheme, PhD dissertation, Stanford University, 2009 http://crypto.stanford.edu/craig
9. VAN EMDE BOAS, P. "Another NP-complete partition problem and the complexity of computing short vectors in a lattice," Technical report 81-04, University of Amsterdam, 1981.
10. AJTAI, MIKLÓS, The shortest vector problem in L2 is NP-hard for randomized reductions, Proceedings of the Thirtieth annual ACM symposium on Theory of computing, May 24-26, 1998, pp. 10-19
11. GOLDREICH, O., MICCIANCIO, D., SAFRA, S. SEIFERT, J.-P.et al., "Approximating shortest lattice vectors is not harder than approximating closest lattice vectors," Inf. Process. Lett. 71, no. 2 (1999): 55–61.
12. LENSTRA, A. K.; LENSTRA, H. W., Jr.; LOVÁSZ, L. Factoring polynomials with rational coefficients. Mathematische Annalen 261 (4): (1982) pp. 515–534.
13. GOLDREICH, O.; GOLDWASSER, S. and HALEVI, S. Public-key cryptosystems from lattice reduction problems, Proceedings of the 17th Annual International Cryptology Conference (CRYPTO 97), Santa Barbara, CA, August, 1997. Also in volume 1294 of Lecture Notes in Computer Science, pp. 112-131, Springer, 1997.
14. HOFFSTEIN, J.; HOWGRAVE-G. N.; PIPHER, J.; WHYTE, W.: Choosing NTRUEncrypt Parameters in Light of Combined Lattice Reduction and MITM Approaches. ACNS 2009: pp. 437-455
15. HOFFSTEIN, J.; HOWGRAVE-G. N.; PIPHER, J.; SILVERMAN, J.H.; WHYTE, W. NTRUSIGN: Digital Signatures Using the NTRU Lattice. CT-RSA 2003: pp. 122-140

Analysis of a Steganographic Method for Hiding an Image into Cover Images Using MATLAB and Image Processing Toolbox

Adriana Borodzhieva[1]

[1] Ruse University "Angel Kanchev", Telecommunications Department, Studentska Str. 8
7017 Ruse, Bulgaria
aborodjieva@ecs.uni-ruse.bg

Abstract. The paper presents an algorithm based on the least-significant bit method for hiding a 24-bit color image into another 24-bit color image in the same format and the same size. The algorithm is implemented using MATLAB and Image Processing Toolbox as the user might choose from a menu between the following formats: bmp, gif and jpeg formats. Applying this method with one cover image the extracted image cannot be reconstructed exactly, i.e. there exists a loss of information. Then the same algorithm is applied for the hidden image to be reconstructed with an accuracy of 100 % when four cover images are used. The results will be used in the educational process in the course "Telecommunication Security", included as compulsory in the curriculum of the specialty "Telecommunication Systems" for the Bachelor degree.

Keywords: Least-significant bit, Steganography, Cryptography, MATLAB.

1 Introduction

Steganography is the art and science of writing hidden messages in such a way that no one, apart from the sender and the intended recipient, suspects the existence of the message. Generally, messages will appear to be something else: images, shopping lists, or some other cover-text. The advantage of steganography over cryptography is that messages do not attract attention to themselves. Whereas cryptography protects the contents of a message, steganography protects both messages and communicating parties [1].

There are a large number of steganographic methods ranging from invisible ink and microdots to secreting a hidden message in the second letter of each word of a large body of text and spread spectrum radio communication. With computers and networks, there are many other ways of hiding information, such as: covert channels; hidden text within Web pages; hiding files in "plain sight"; null ciphers. Today, steganography allows a user to hide large amounts of information within image and audio files. These forms of steganography often are used in conjunction with cryptography so that the information is doubly protected; first it is encrypted and then hidden so that an adversary has to first find the information and then decrypt it [2].

A large number of commercial steganographic programs use the Least Significant Bit (LSB) embedding as the method of choice for message hiding in 24-bit, 8-bit color images, and grayscale images. It is commonly believed that changes to the LSBs of colors cannot be detected due to noise that is always present in digital images [2].

2 An Algorithm for Hiding an Image into another Image and Extracting the Hidden Image

An algorithm for hiding an image (*hidden*, fighter F16) into another image (*COVER*, sunrise in San Clemente) and extracting the *hidden* image using the program and computing environment MATLAB [4] and its extension Image Processing Toolbox is investigated and presented in the paper. Both images must have the same size (in this case it is 310 x 516, after artificially "cutting off" the *COVER* image to the needed size). The processed images are both of type BMP using 24-bit depth of the colors [3].

The algorithm for hiding the *hidden* image into the *COVER* image is illustrated in Fig. 1. Each pixel (*I, J*) of the *COVER* image is presented by 8-bit sequences of red, green and blue components, which bits are marked as $R1 \ldots R8$, $G1 \ldots G8$ and $B1 \ldots B8$ respectively. Each pixel (*I, J*) of the *hidden* image is also presented by 8-bit sequences of red, green and blue components, which bits are marked as $r1 \ldots r8$, $g1 \ldots g8$ and $b1 \ldots b8$ respectively. The hiding process consists of the following stages [3]:

1. The 24-bit color image of fighter F16 subject to hiding is stored in the array *hidden*, and the 24-bit color image of sunrise in San Clemente served as carrying media of the *hidden* image is stored in the array *COVER*. Both arrays are with size of 310 x 516 x 3 (3-dimensional for color images) and contain 8-bit unsigned integers.

2. The number *n*, $1 \le n \le 7$, is chosen. It shows the number of the least significant bits of the *COVER* image, that will be used for hiding the most-significant bits of the *hidden* image; if $n = 8$ is chosen, then all the eight bits of the *COVER* image are replaced with all the bits of the *hidden* image, as the *STEGO* image, containing the results from the hiding process, is completely identical with the *hidden* image.

3. The variable *BC* is formed. It contains the result from running the function *bitcmp* (bitwise complement) of the variable $Pn = 2^n - 1, n = 1 \div 8$ to 8-bit unsigned integer, i.e. it results in an 8-bit sequence of 1s where the last *n* 1s are replaced by 0s (Table 1). The values of *n*, greater than or equal to 9 are not applicable in the case, since more than 8 bits are necessary for representing the number *Pn* in a binary form.

4. Setting to 0s the least significant *n* bits of the three color components for each pixel of the *COVER* image by running the operation *bitand* (bitwise AND) of the operands *COVER* and *BC*, as the variable *BA* is created.

5. Setting to 0s the most significant $8 - n$ bits of the three color components for each pixel of the *hidden* image by running the operation *bitshift* (bitwise shift) of the operand *hidden* to $n - 8$ bits, i.e. to $8 - n$ bits to the right. As a result the variable *BS* is created.

6. Creating the variable *BO* as a result of running the operation *bitor* (bitwise OR) of the operands *BA* and *BS*. The variable *BO* represents the *STEGO* image that contains the image of the sunrise together with the *hidden* image of the fighter, *STEGO* = *BO*.

Table 10. Running the operation *bitcmp*

n	1	2	3	4	5	6	7	8	9
Pn = 2^n - 1	1	3	7	15	31	63	127	255	511
BC - DEC	254	252	248	240	224	192	128	0	-
BC - BIN	11111110	11111100	11111000	11110000	11100000	11000000	10000000	00000000	-

The extracting process of the *hidden* image from *STEGO* image includes running the operation *bitshift* (bitwise shift) of the operand *STEGO* in $8 - n$ bits to the left. In the

extracting process the least significant $8-n$ bits of the *hidden* image are lost and the quality of the extracted image is not the same as the original image.

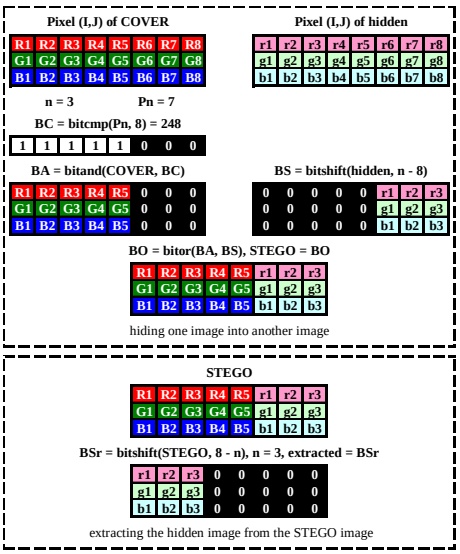


Fig. 1. An algorithm for hiding one image into another image and extracting the *hidden* image from the *STEGO* image

The results from running the script in MATLAB are illustrated in the figures below. The original images – sunrise in San Clemente (*COVER* image) and fighter F16 (*hidden* image) are presented in Fig. 2. The corresponding *STEGO* images *STEGO i* $(i=1\div 7)$ with the hidden image of the fighter and the extracted images *extracted i* are presented in Fig. 3 ... Fig. 9 when the least significant *i* bits of each color components (*R*, *G*, *B*) for each pixel of the *COVER* image are used.

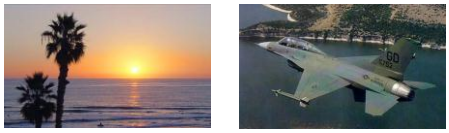


Fig. 2. *COVER* and *hidden* images

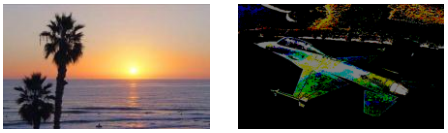


Fig. 3. *STEGO* 1 and *extracted* 1 images

Fig. 4. *STEGO* 2 and *extracted* 2 images

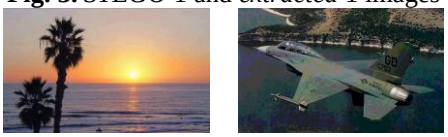


Fig. 5. *STEGO* 3 and *extracted* 3 images

Fig. 6. *STEGO* 4 and *extracted* 4 images

Fig. 7. *STEGO* 5 and *extracted* 5 images


Fig. 8. *STEGO* 6 and *extracted* 6 images


Fig. 9. *STEGO* 7 and *extracted* 7 images

3 An Algorithm for Hiding an Image into other Four Images and Extracting the Hidden Image

An algorithm for hiding an image in BMP format into other four images in the same format and extracting the *hidden* image is implemented using fourfold applying the algorithm described in Section 2 as four *COVER* "carrier" images (*COVER* 1, *COVER* 2, *COVER* 3 and *COVER* 4) are used for hiding the *hidden* image. All the five images must have the same size, in the case 640 x 480 pixels. The processed images are in BMP format with 24-bit depth of the colors. The algorithm for hiding the *hidden* image into the four images *COVER* 1, *COVER* 2, *COVER* 3 and *COVER* 4 is illustrated in Fig. 10. All pixels (*I, J*) of the images *COVER* 1, *COVER* 2, *COVER* 3, *COVER* 4 and *hidden* are presented by 8-bit sequences of red, green and blue components. The hiding process contains the following stages:

1. 24-bit color image of the fighter F16 subject to hiding is stored in the array *hidden*. 24-bit color "carrier" images of San Clemente, Italy, Germany and Scotland are stored in the arrays *COVER* 1, *COVER* 2, *COVER* 3 and *COVER* 4. These four images will serve as carrier media for the hidden data. All five arrays are with size of 640 x 480 x 3 (3-dimensional for color images) and contain 8-bit unsigned integers.

2. The variable *BC* = *bitcmp*(2,8) is formed. It contains the result from running the operation *bitcmp* (bitwise complement) of the variable $Pn = 2^n - 1$, $n = 2$, to 8-bit unsigned integer, i.e. as a result of running the operation *bitcmp* an 8-bit sequence of 1s is obtained where the last bits are replaced by 0s (Table 1). Here, the special feature is that each of the "carrier" images will carry only 2 bits of the *hidden* image, i.e. *COVER* 1 will hide the most significant two bits of the *hidden* image, *COVER* 2 – the next two bits of the *hidden* image, *COVER* 3 – the next two bits, and *COVER* 4 – the last two, the least significant two bits of the *hidden* image.

3. Setting to 0s the least significant $n = 2$ bits of the three color components for each pixel of all four images *COVER i* by running the operation *bitand* (bitwise AND) of the operands *COVER i* and *BC*, as the variables *BA i*, $i = 1 \div 4$ are created.

4. The new variables *BS i*, $i = 1 \div 4$, are created by running the operation *bitshift* (bitwise shift) of the operand *hidden*. The shift of the operand *hidden* is correspondingly to 6 bits (when $i = 1$), 4 bits (when $i = 2$), 2 bits (when $i = 3$) и 0 bits (when $i = 4$). The most significant two bits of *hidden* are in the positions of the last two least significant bits of *COVER* 1: *BS*1 = *bitshift*(*hidden*, −6). Respectively, the next two bits of *hidden* are in the positions of the last two least significant bits of *COVER* 2, i.e. *BS*2 = *bitshift*(*hidden*, −4). The variables *BS*3 and *BS*4 are created by the instructions respectively: *BS*3 = *bitshift*(*hidden*, −2), *BS*4 = *bitshift*(*hidden*, 0).

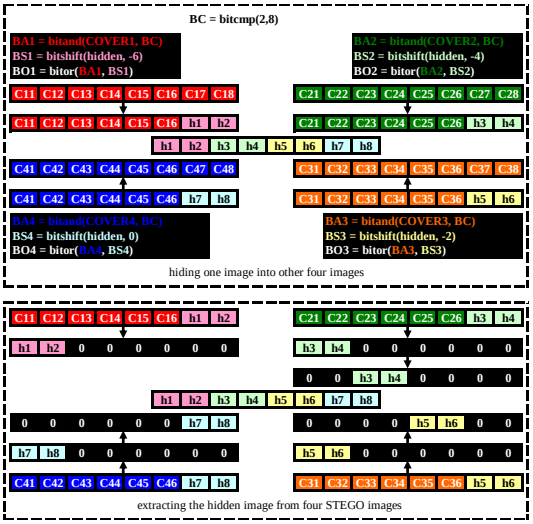


Fig. 10. An algorithm for hiding one image into four other images and extracting the *hidden* image from four *STEGO* images

5. Creating the variables *BO i*, $i = 1 \div 4$ as a result of running the operation *bitor* (bitwise OR) of the operands *BA i* and *BS i*. The variables *BO i* represent the *STEGO* images, i.e. the images of San Clemente, Italy, Germany, Scotland together with the information hidden in them (bits form the *hidden* image of the fighter). Therefore, *STEGO i = BO i*, $i = 1 \div 4$.

When reconstructing the *hidden* image, the extracting process is also applied four times as two bits from the *hidden* image are extracted from each of these four *STEGO* images.

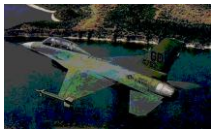
The four *COVER* images *COVER*1 (San Clemente), *COVER*2 (Edinburgh castle), *COVER*3 (Italy, Capri) and *COVER*4 (Germany) are given below (Fig. 11 … Fig. 14). The *STEGO* images and the *extracted* images as a result from running the scripts in MATLAB are presented also (Fig. 11 … Fig. 14).

At the end, the 100 % reconstructed image is given. The original *hidden* image and the *extracted* image are presented in Fig. 15.

The scripts implemented in MATLAB and Image Signal Processing [4] will be used in the educational process in the course "Telecommunication Security" and it will enable students to comprehend the principles of steganography based on the LSB-method. Future work will include creating a friendly graphical user interface for simulating the processes of hiding an image into a cover image or into four other images as well as hiding a text or an image into music files.

*COVER*1 (San Clemente)

*STEGO*1

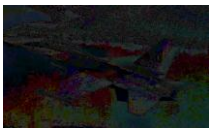


*extracted*1

Fig. 11. *COVER*1, *STEGO*1 and *extracted*1 images

*COVER*2 (Edinburgh)

*STEGO*2

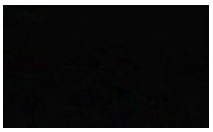


*extracted*2

Fig. 12. *COVER*2, *STEGO*2 and *extracted*2 images

*COVER*3 (Italy, Capri)

*STEGO*3



*extracted*3

Fig. 13. *COVER*3, *STEGO*3 and *extracted*3 images

*COVER*4 (Germany)

*STEGO*4

*extracted*4

Fig. 14. *COVER*4, *STEGO*4 and *extracted*4 images

Fig. 15. The original *hidden* image and the *extracted* image

CONCLUSION

LSB is one of the easiest steganography methods for program implementation, providing relatively good quality of the recovered *hidden* image. An algorithm for hiding a color image into another color image in the same format, using 24-bit color depth, is presented in the paper. The algorithm is implemented using MATLAB. The extraction of the *hidden* image, however, losses a part of information contained in the least significant bits, which can be visible by the human eye in some cases. There are a number of areas such as aeronautics, weather systems or military missions requiring the *hidden* image to be recovered 100 %. This objective is achieved by applying fourfold the algorithm proposed in the paper using four "carrier" images, each of them carries two bits of the *hidden* image. In the recovery of the *hidden* image, the process of extraction is also applied fourfold: two bits of the *hidden* image are extracted from each of the four received *STEGO* images.

References

1. http://en.wikipedia.org/wiki/Stegano
2. http://www.garykessler.net/library/steganography.html
3. http://conf.uni-ruse.bg/bg/docs/sns/2009/SNS2009_EEA.pdf
4. http://www.mathworks.com

Software Tool for Implementing Encryption and Decryption Processes Using Classical Ciphers

Adriana Borodzhieva[1]

[1] Ruse University "Angel Kanchev", Telecommunications Department, Studentska Str. 8
7017 Ruse, Bulgaria
aborodjieva@ecs.uni-ruse.bg

Abstract. Nowadays cryptography is applied for encrypting information in computer-communication systems. There exist many cryptographic algorithms. The paper presents a software tool for implementing encryption and decryption processes using some classical ciphers, as Playfair cipher or Playfair square, two-square ciphers and four-square cipher. These algorithms are modified with the aim of using Bulgarian texts for encryption/decryption. The symmetric encryption technique encrypts pairs of letters (digraphs), instead of single letters as in the simple substitution cipher and rather more complex Vigenère cipher systems then in use. Frequency analysis can still be undertaken, but on the 600 possible digraphs rather than the 26 possible monographs. The frequency analysis of digraphs is possible, but considerably more difficult – and it generally requires a much larger cipher-text in order to be useful. Design and operation of a software tool for implementing encryption and decryption processes using classical cipher are described in the paper. The architecture of the system and modules for generating keys and encrypting/decrypting are presented. The software tool is developed in MATLAB and it will be applicable to the educational process in the course "Telecommunication Security".

Keywords: Cryptography, cryptanalysis, classical ciphers, MATLAB.

1 Introduction

The Playfair cipher is a symmetric encryption technique, the first literal digraph substitution cipher. The scheme was invented in 1854 by the noted British scientist Sir Charles Wheatstone (1802 – 1875), but the usage of the cipher was promoted by Lord Lyon Playfair (1818 – 1898) and today it is known as the Playfair cipher (or Playfair square) [1, 2, 3]. The four-square ciphers and the two-square ciphers are also symmetric encryption techniques. The four-square cipher was invented by the famous French cryptographer Felix Delastelle [4], and the two-square cipher (also called double Playfair) was developed to ease the hard nature of the large encryption/decryption matrix used in the four-square cipher while still being slightly stronger than the (single-square) Playfair cipher [5].

The technique used in all these algorithms encrypts pairs of letters (digraphs), and thus falls into a category of ciphers known as polygraphic substitution ciphers. This adds significant strength to the encryption when compared with monographic substitution ciphers which operate on single characters. The use of digraphs makes the two-square/four-square technique less susceptible to frequency analysis attacks, as the analysis must be done on 676 possible digraphs rather than just 26 for monographic substitution. The frequency analysis of

digraphs is possible, but more difficult - and it generally requires a much larger cipher-text in order to be useful [4, 5].

The principles of encryption/decryption using the Playfair cipher, four-square ciphers and two-square ciphers are presented in details in [1, 3, 4, 5].

2 Software Tool for Encryption and Decryption Processes Using Classical Cipher

Design and operation of a software tool for implementing encryption and decryption processes using Playfair cipher, four-square cipher and two-square ciphers are described in the paper. The architecture of the system and modules for generating the keys and encrypting/decrypting are presented. The software tool is developed in MATLAB [6] and it will be applicable to the educational process in the course "Telecommunication Security".

The software tool allows the user to make his/her choice of language, type of operation and cipher used (Fig. 1). Through menu (Fig. 1, a) the user selects the language used for the interface and the input text for encryption/decryption (English or Bulgarian). Again through menu (Fig. 1, b) the user selects the type of the operation (encryption or decryption, the texts of messages are determined depending on the choice of language). Finally through menu (Fig. 1, c) the user selects a cipher to be used in the operation (encryption or decryption), and he/she can choose one of the following ciphers: Playfair, four-square or two-square ciphers.

The software tool consists of three modules:

1) Module 1 – Generating the key;
2) Module 2 – Encrypting;
3) Module 3 – Decrypting.

The architecture of the system and the structures of the modules for generating the key, encrypting and decrypting are presented in Fig. 2.

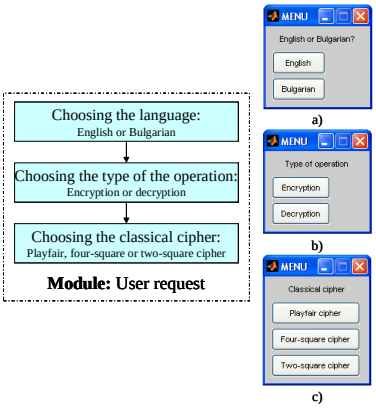


Fig. 1. Processing user requests

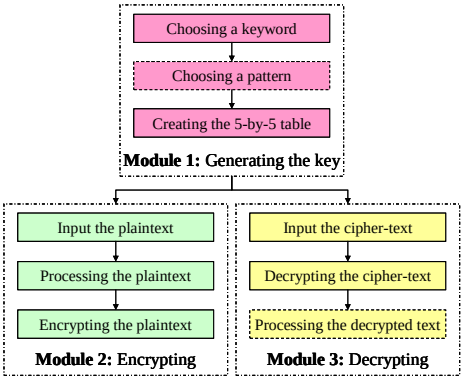


Fig. 2. Architecture of the system

Module 1 – Generating the Key

First, a keyword or phrase for building 5-by-5 table is chosen by the user. For texts in Bulgarian 6-by-5 table is applied (Fig. 3). In [1] the algorithm for generating the key is presented in details. Here, the user can choose from 11 key words. Then 5-by-5 (for English) and 6-by-5 (for Bulgarian) tables are formed and a picture of the square to be used is presented in a graphical window (for example when the key word CRYPOGRAPHY or КРИПТОГРАФИЯ is chosen).

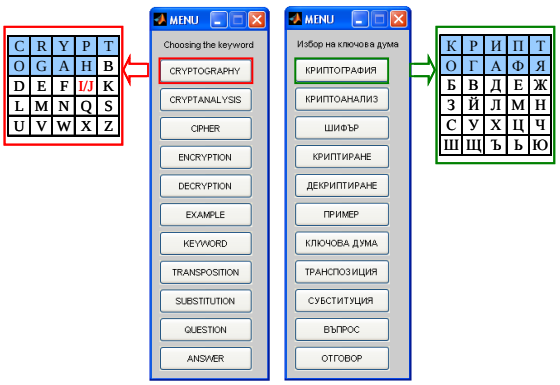


Fig. 3. Module "Generating the Key"

Module 2 – Encrypting

The plaintext for encrypting is entered by the user. It is stored in the string variable S. Next the plaintext is processed in three stages: 1) transforming all letters "J" into "I" (Processing 1, only for English texts, Fig. 4); 2) dividing the plaintext in digraphs and eliminating the same letters in each digraph (Processing 2, only for Playfair cipher, Fig. 5) – if the letters in the digraph are 'ZZ' (in English) / 'ЬЬ' (in Bulgarian) then the low-frequency letter 'X' / 'Ъ' is added between them, otherwise the low-frequency letter 'Z' / 'Ь' is added; 3) if the number of the letters in the plaintext (after modifying the text according to Step 2) is odd, then the plaintext is complimented by the low-frequency letter – 'X' / 'Ъ', if the last letter is 'Z' / 'Ь', and otherwise by 'Z' / 'Ь', to satisfy the condition in Step 2 (Processing 3, Fig. 6).

Afterwards the encryption of the plaintext is realized applying the rules of the Playfair cipher or four-square cipher, or two-square cipher. The process of encrypting using Playfair cipher is presented in Fig. 7 [1], four-square cipher – in Fig. 8, two-square ciphers – in Fig. 9. All the operations in these figures are fulfilled for each i = 1 : 2 : Ls – 1. The cipher-text is stored in the variable Sc. The variable is initially set as an empty string and it is changed during the encryption process. First the coordinates [xa, ya] and [xb, yb] of the symbols S(i) and S(i + 1) in the 5-by-5 table (for English texts) or 6-by-5 table (for Bulgarian texts) are determined. Next the "new" coordinates [xan, yan] and [xbn, ybn] of the encrypted symbols are calculated depending on the location of the symbols S(i) and S(i + 1) in the table/s. The differences for Bulgarian texts are given in the figures.

For English texts: if xa = xb, i.e. the symbols in the plaintext are in the same row of the 5-by-5 table, then xan = xa, xbn = xb, yan = ya + 1 and ybn = yb + 1 (Fig. 7). It means that the letters in the digraph are replaced with the letters to their immediate right respectively. If yan = 6, i.e. the first letter in the original pair is on the right side of the row, then yan = 1, and the letter in the plaintext is replaced with the letter of the left side of the row. By analogy, if ybn = 6, i.e. the second letter in the original pair is on the right side of the row, then ybn = 1, and the letter in the plaintext is replaced with the letter of the left side of the

row. Similar operations are done if the letters of the original pair are in the same column (ya = yb).

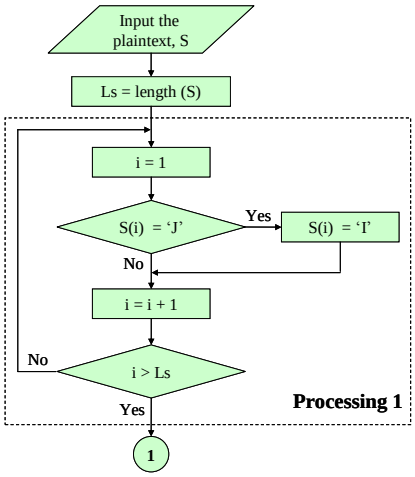


Fig. 4. Processing 1 of the plaintext (only for English texts)

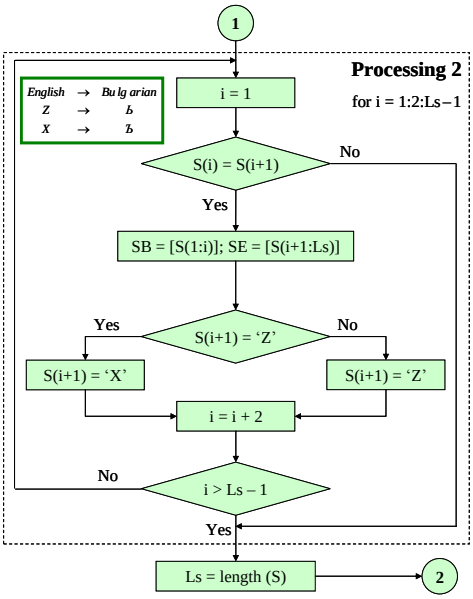


Fig. 5. Processing 2 of the plaintext (only for Playfair cipher)

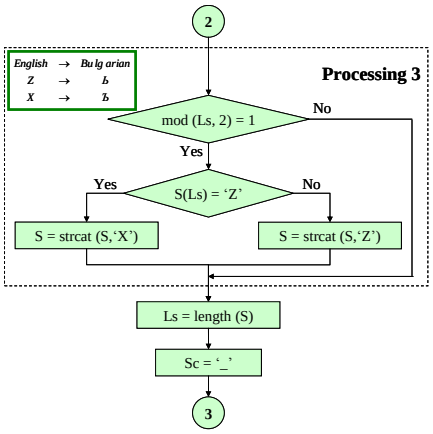


Fig. 6. Processing 3 of the plaintext

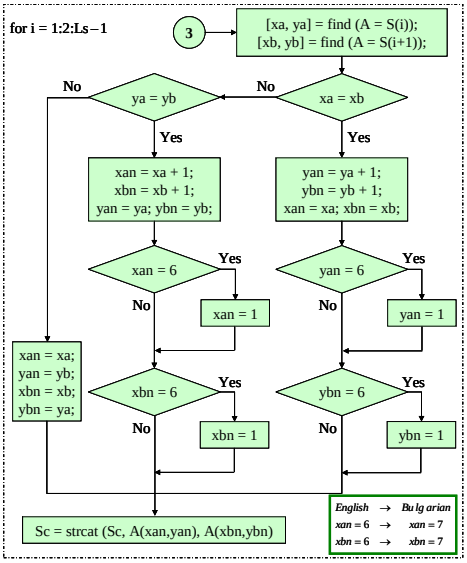


Fig. 7. Encrypting the plaintext (Playfair cipher)

Otherwise, when the letters are not in the same row or column, the "new" coordinates of the letters in the cipher-text are calculated as follows: xan = xa, xbn = xb, yan = yb and ybn = ya. It means that the letters in the original pair are replaced with the letters in the same

row respectively but at the other pair of corners of the rectangle defined by the original pair. The first encrypted letter of the pair is the one that lies in the same row as the first plaintext letter. For four-square cipher (Fig. 8) and two-square ciphers (Fig. 9) the "new" coordinates [xan, yan] and [xbn, ybn] of the encrypted symbols are calculated by analogy.

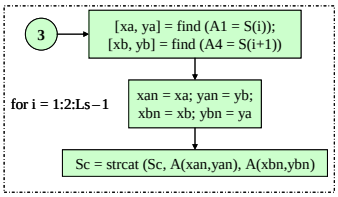


Fig. 8. Encrypting the plaintext (four-square cipher)

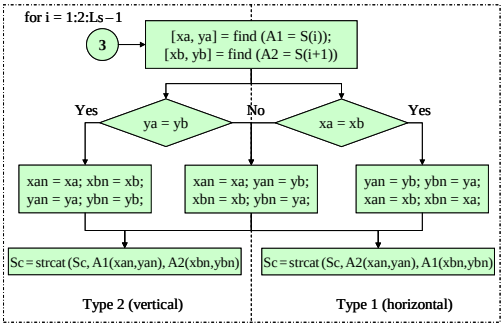


Fig. 9. Encrypting the plaintext (two-square cipher)

Module 3 – Decrypting

Decrypting the messages is implemented applying the inverse of the rules used in the encrypting process. At this stage, additional processing of the encrypted text (dropping any extra "X"s, "Z"s or "Q"s) that do not make sense in the final message after finishing) is not realized.

Fig. 10. Principle of operation of Playfair ciphers

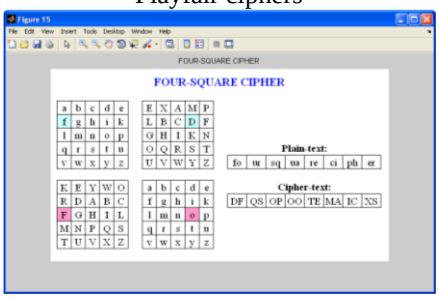


Fig. 11. Principle of operation of four-square ciphers

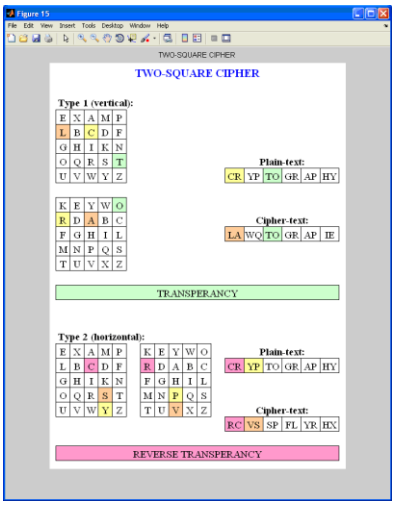


Fig. 12. Principle of operation of two-square ciphers

The software tool allows the principle of operation of the cipher applied in the process of encryption to be displayed in a graphic window for easier assimilation by the students. A picture illustrating the principle of operation of Playfair cipher is shown in Fig. 10. Similar images are displayed when four-square or two-square ciphers are used (Fig. 11 and Fig. 12).

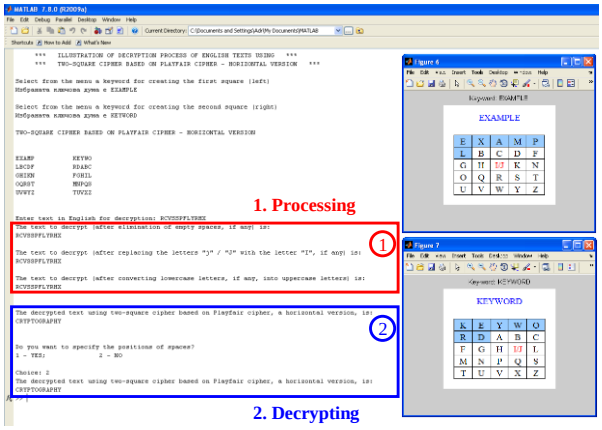


Fig. 13. A screenshot of running the software tool for decrypting the text RCVSSPFLYRHX

A screenshot of running the software tool is given (Fig. 13) as the English text RCVSSPFLYRHX is first processed (for encryption or decryption, spaces must be eliminated) and then decrypted using two-square horizontal cipher based on Playfair cipher, with 5-by-5 tables built on the basis of key words EXAMPLE and KEYWORD. The decrypted text is CRYPTOGRAPHY. The result is the same as in Fig. 12.

CONCLUSION

Design and operation of a software tool for implementing encryption and decryption processes using Playfair cipher, four-square and two-square ciphers are described in the paper. The architecture of the system and modules for generating the keys and encrypting/decrypting are presented. The software tool is developed in MATLAB. It will be applicable to the educational process in the course "Telecommunication Security" and it will enable students to comprehend the principles of encrypting and decrypting using these ciphers. Future work will include creating a friendly graphical user interface (GUI) for simulating the processes of encrypting and decrypting using different ciphers besides these ciphers – for example, Polybius square, ADFG(V)X, Bifid, Trifid, etc. The simulation results will be compared to the theoretical ones.

References

1. Borodzhieva, A., P. Manoilov. Software Tool for Implementing Encryption and Decryption Processes Using Playfair Cipher. International Symposium for Design and Technology of Electronic Packaging, 14th Edition, SIITME 2008, September 18 – 21, Predeal, Romania, Conference proceedings, pp. 140 – 144, ISSN 1843-5122.
2. http://en.wikipedia.org/wiki/Classical_cipher
3. http://en.wikipedia.org/wiki/Playfair_cipher
4. http://en.wikipedia.org/wiki/Four-square_cipher
5. http://en.wikipedia.org/wiki/Two-square_cipher
6. http://www.mathworks.com/

Technology-Enhanced Personalized Education

**Valentina Terzieva, Petia Kademova-Katzarova,
Katia Todorova, Rumen Andreev**

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences Acad. G. Bonchev Str., Bl. 2, 1113 Sofia, Bulgaria
{valia, petia, katia, rumen}@isdip.bas.bg

Abstract. Application of computers and information technologies has radically changed technology of learning. It is important how effectively it is used. The full potential of technology-enhanced learning is revealed in delivering personalized and individualized learning. Clearly, it is not just a technology innovation that drives development of e-learning. People need to get new information and knowledge that are fast changing. For everyone who needs education the most effective way is teaching on demand, but that is expensive. To exploit ICT capabilities for building and delivering aforesaid personalized and individualized learning is a cost-effective way to achieve better education. The paper presents an overview of computer-aided teaching methods used for adaptation and personalization of learning resources according to the student's psychology and cognitive profile, their interest and learning background. In addition tools and techniques for adaptation to user's preferences for appearance of learning environment and representation of learning content are mentioned.

Keywords: e-learning, personalization, adaptation, learning style

Technology-enhanced Learning

Technology-enhanced learning has aroused intense interest during last decade because of vast spectrum of functionality that ICT offer. They support not only richness of learning resource (including text and multimedia), various presentation modes, but personalized learning tailored to the unique requirements, preferences and knowledge background of a learner. Technology-enhanced learning facilitates a major shift towards more knowledge and information driven educational systems that will make them more flexible and adaptive. Recently different kinds of realization of technology-enhanced e-learning systems such as intelligent tutoring systems, personal learning environments and Web-based communities of practice become wide spread.

The educational process can be realized in different ways, by different means and in various environments in correspondence with the concepts of adopted teaching model. Technology-enhanced learning is result of usage information and communication technologies (ICT) and especially computers in the realization of the basic components of an education system. There are several technology-based approaches for supporting the educational activities. In the framework of the cognitive paradigm based on an information processing loop including the teacher(s) and learner(s), the computerization of teaching activity results in development of the following types of systems:

- E-learning systems for blended learning – traditional instruction-led teaching supplemented with a various technology-based pedagogical methods; it conceives of the idea that the teaching is a process of knowledge transferring;
- Learning management systems (LMS) – they support not only various teaching functions like authoring, knowledge management, certification and training, mentoring, personalization, discussion boards, etc., but also some administrative activities concerning organization of education;
- Adaptive Web-Based Educational Systems [1, 2] – they are basis of the emerging Educational Semantic Web that realizes one of the constructivist approaches to fulfilment of teaching activity known as "communities of practice" – it guarantees sharing of learning resources that are products of teaching activity and achievement of interoperability among various teaching activities;
- Computer-supported learning design systems [3] – they support the design of units of learning that ensure the realization of a teaching-learning (information) process.

The traditional education mainly supports an information-oriented view on learning. This perspective imposes an instructional approach to knowledge delivery. Instruction can be viewed as a function of human communications, which key elements are messages and symbols. Since they transfer knowledge and meanings, a teaching-learning process focuses on them [4]. Modern education can not only be surmounted with the possibilities of communication-based paradigm. Contemporary ICT provide tools for a significant shift in paradigm towards technology-enhanced personalized education. Technology-enhanced learning has unique attribute that attracts serious interest – it can support an individual learning customized to the learner's characteristics, preferences, prior knowledge and competences, etc. However, the development of such personalized environment is a big challenge, as it should provide different learning course for each individual user and/or different context. The principles of development of a technology-enhanced e-learning environment have to guarantee realization of the environment's attributes such as personalization, users' collaboration, etc. The basic principles can be summarized as follows:

- Supporting determination of individual cognitive and psychological characteristics of a learner (model of student);
- Technology-oriented tools supporting realization of different teaching methods (effective education);
- Provision with appropriate tools to assist professionals in creation and personalization of learning resources (authoring tools);
- Communication tools supporting cooperation and collaboration, knowledge and good practice sharing;
- Technology-based tools for storage and reuse of data, resources, facilities, etc;
- Learning process adaptation to individual learning abilities of the students;
- User-friendly ICT-based environment for successful teaching/ learning (usability).

The above mentioned basic principles are guiding rules in development of a personalized e-learning environment. The main components of the environment and relationships among them reveal how these principles are built in [5]. The functional components that realize the basic principles are presented by Student's Model, Model of Teaching, Reuse Strategy, Authoring Model and Model of Knowledge Domain, (Fig. 1). This is a model-based framework ("meta-framework") that describes the architecture of a development system, which realizes the upper development principles. The presented system organization can serve as a basis for construction of technology-enhanced e-learning environment that support personalized education.

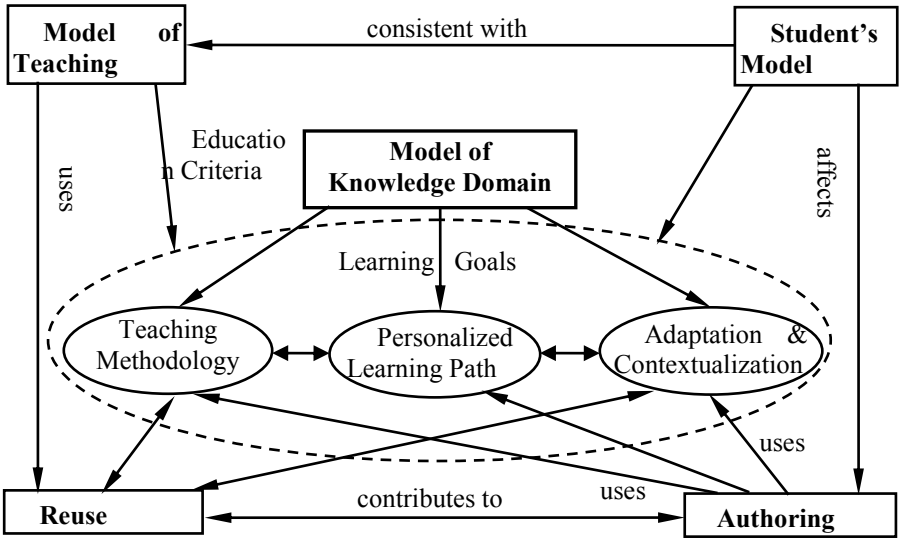


Fig. 1. Framework of personalized learning environment

The Model of Teaching describes selected teaching method in correspondence with the learning goals and cognition characteristics of the Student's Model. It ensures information that affects an authoring process. For each student his/her individual model is composed of psychological and cognitive characteristics, knowledge background and competences. Cognitive characteristics include learning abilities and learning style of a student. They are determined by assessment of student's learning preferences and psychological tests. The learning style presents cognitive forms that a student uses. It concerns also the way of knowledge construction. The assessment of competence stores information about student's achievements at different stages of educational process.

The Model of Knowledge Domain represents a field of knowledge, which have to be acquired by teaching/learning activities. This is the area where learning goals are determined. Since the learning environment is developed to support achievement of these goals, the domain model presents its rationale, i.e. its logical basis. The Education Criteria are reflected in the process of constructing the Model of Knowledge Domain.

The education is presented as a teaching-learning process. The teaching element of this process is responsible for production of educational materials necessary for successful learning. They are learning resources provided by the e-learning environment, which have to

supply learning activities and support an adopted teaching approach. The process of producing of educational resources passes through three stages presented by Teaching Methodology, Personalized Learning Path and Adaptation and Contextualization.

A Teaching Methodology is a set of rules that prescribes how the student can achieve his personal learning goals in a certain context in the most effective way. The Teaching Methodology defines the way of knowledge presentation that is in agreement with student's learning style. An application of a Teaching Methodology for a specific learning goal and a specific context specifies a learning process – it defines a Personalized Learning Path. The learning resources are context-dependent. The module Adaptation and Contextualization serves for representation of learning resources in specific context, and adaptation to the student's requirements and preferences. It presents an adaptation strategy, which determines how and what to adapt/ personalize based on the information stored in the Model of Student, Personalized Learning Path, and Teaching Methodology. The Personalized Learning Path represents a sequence of learning activities that is in accordance with learner's cognitive style, i.e. it ensures effective acquiring of knowledge.

The Authoring Model represents the management of the creation of learning resources and supports the authors' collaboration. The collaboration form depends on the Reuse Strategy applied in the environment. The reuse method determines the description modes of learning resources and applied teaching methodology. It also maintains an approach for their storage and search. It supports multiple uses of resources developed by different authors in the e-learning environment.

Personalized Learning

Learning environment can be organized in different ways. Depending on its degree of customization for the learner we can discern four general types (Fig. 2):

- Classic – traditional full training class;
- Training groups – within the class are formed groups of students with similar abilities and interests;
- Personalized learning – learning is fully consistent with the capabilities and characteristics of the individual student;
- Learning process is almost completely controlled and directed by the learner.

In the classical case teaching, learning and testing of knowledge and skills are performed according to the government standards and requirements for each level, regardless of the specific features of the individual learners. The teacher is confined in the choice of applied tools and methods, for he has to ensure a "medium level" that covers wide range of students. This level is usually too high for children with special educational needs or other problems, while insufficient for children with higher skills than average.

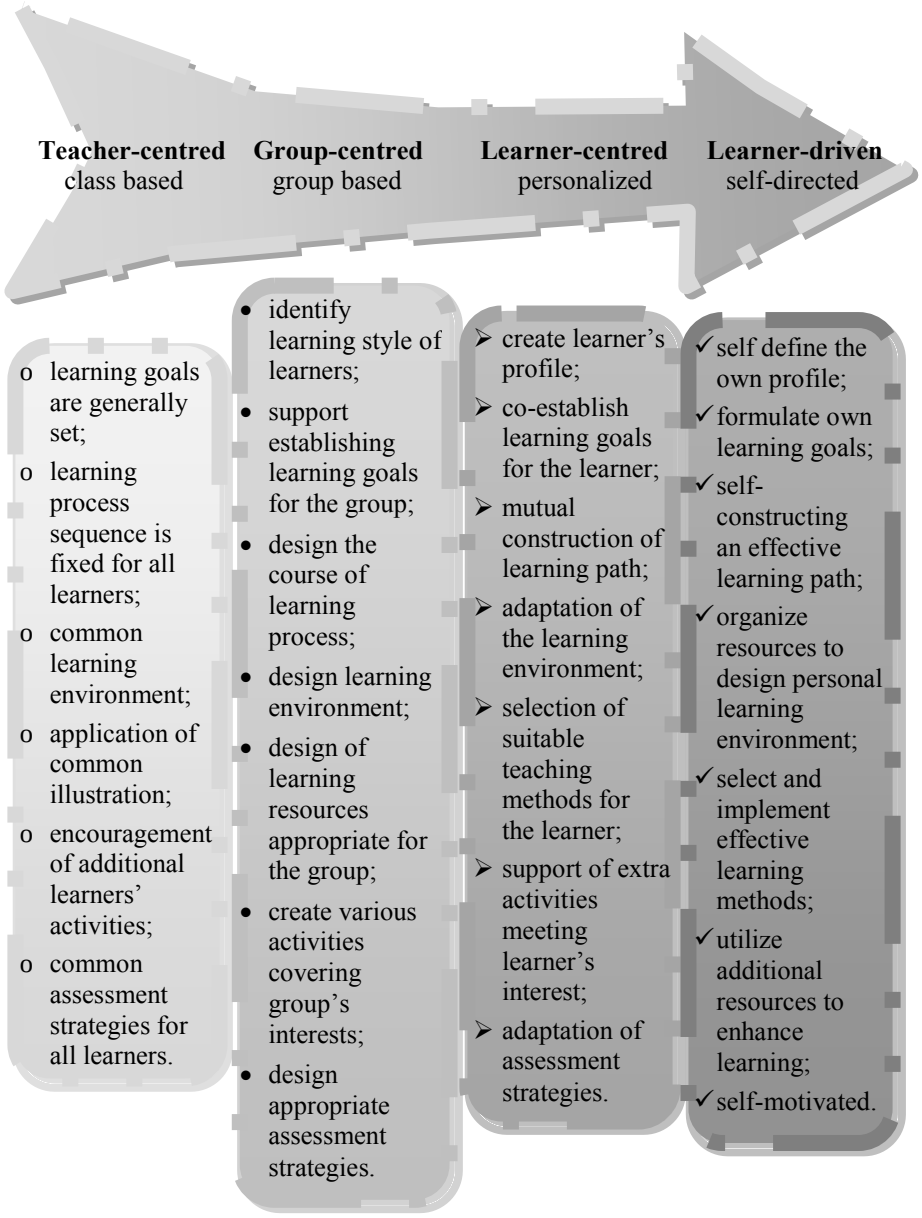


Fig. 2. Evolution of Education

Through additional activities (presentations, visits to sites related to the learned subject etc.) better results can be achieved, but the general disadvantage of this type of learning (the class is treated as one) is not removed. "Teacher is introduced to competency-based learning. Learning is still part of a standards-driven, time-based grade level system" [6].

When working in groups a differentiated approach and means can be applied to comply with the level of each group. For example, the teaching material can be presented in a different sequence (shifting of some lessons), can be illustrated in a different way and can be used for different tasks, exercises and different types of exams. Additional activities in support of the training may take the form of workshops or projects; each group receives an assignment that causes the most interest, respectively produces higher motivation and better learning curriculum.

Creating a personalized learning path for each student is the best way of training, but requires many additional activities – ranging from the creation of cognitive-psychological profile of each student to the creation of a "tailored" learning process. Again, additional activities may be allocated according to the interests and abilities of the individual student and the teacher can assess what is most beneficial for assimilation of the learning matter.

The last type of learning environment where the learner largely alone creates his personal curriculum actually is feasible only for adult learners who deliberately chose this training and pursue a realization – students, graduates, doctoral students participating in courses, lifelong learning and others.

To achieve a high quality of all above mentioned types of education is necessary to use appropriate technical means. But the implementation of modern ICT for the last two types is indispensable. The personalized learning path compliant with the specific cognitive and psychological characteristics of each learner can be achieved only by means of ICT.

Techniques for Personalization/ Adaptation

In order to apply effective personalized education an adequate model of each student is needed. It will improve the ability of technology-enhanced educational system to understand and take into account users' needs and preferences. The model describes personal characteristics and preferences, personality, learning style etc. It also describes the favourite presentation format, content display, sound and video effects, environment appearance etc. It includes the behavioural patterns (individual manners during learning process) and specifics in the cognitive processes that underlie the student's actions. Furthermore, it supports defining learning style according to the expressed competences, educational background, specific deficits, and learning goals.

The learners undergo a psychological test that tackles their cognitive abilities and psychological characteristics. Their strong and weak features are detected and a model of each student is formed. The essential features acquired for each learning subject are selected to be employed in an effective learning process. The characteristics most important for the setting up of the personal profile concern the cognitive and the psychological aspects of the student's model. This model defines the learning style and indicates suitable pedagogical methods and tools to be applied. The cognitive abilities building the personalized profile include:

- Memorizing (short / long term memory),
- Orientation, coordination,
- Interpreting abstract terms and symbols,
- Attention, concentration,
- Learning capacity,
- Way of thinking – in terms ("sequential") or in pictures ("quasi parallel").

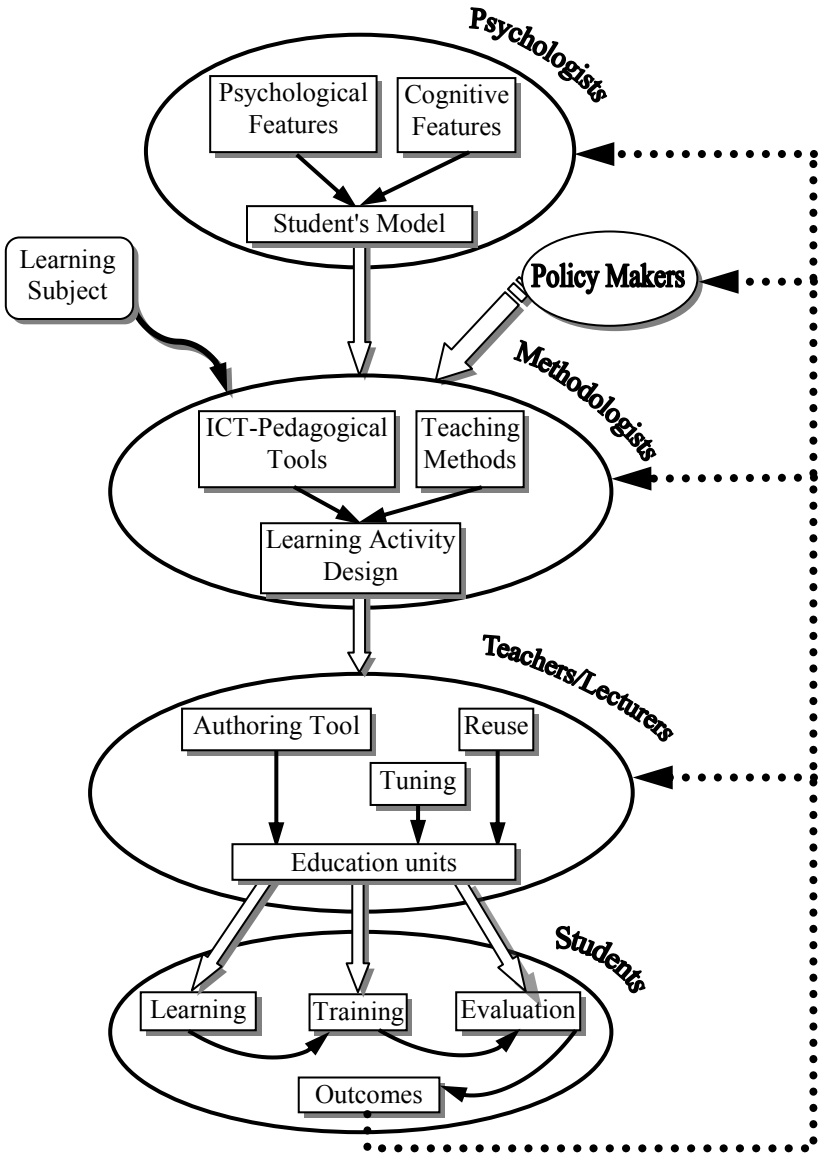


Fig. 3. Process-oriented model of personalized education

The psychological features taken into consideration are: self-assessment and motivation, imagination, patience, emotionality, communication skills.

The student's model defines the appropriate learning style for the student and hence the appropriate teaching methods and pedagogical tools to be chosen. On that basis are designed

the learning activities (Fig. 3). For every student a set of learning activities is chosen that forms a personalized learning path. Every step in the learning path is designed to meet the personal needs of the student. The cognitive profile determines the methods and tools to be applied to accomplish effectively the teaching / learning goals. The educational goals set by government requirements (volume and level of knowledge to be attained) are to be considered too. The learning activities are implemented by the composing of new and adaptation of the reusable education units and their presentation. The personalised unit is prepared using the existing and the composed ones, and it is tuned according to local and language specifics, custom, traditions, presentation format etc. The education units are used not only in the teaching process, but also for training and tests. The outcomes bear information about the efficiency of the whole education process and are used for refinement of the student's model, adjusting the teaching methods and employed pedagogical tools and developing education units, but also (as statistic data) as feedback for policy makers. The individual curriculum and outcomes are stored in data base and used for further update of the student's record to improve the teaching practice. Thus the technology- enhanced environment integrates all the necessary data and tools for the efficient educational process.

ICT for Effective Educational Process

Analyzing widespread educational environments and e-learning systems we suggest framework of personalized learning environment supporting learner centred education. It should be focused on the educational quality that is determined by many factors. One of them is adaptability to the learner's preferences and learning style. Some of the others are proper evaluation of students' needs and characteristics, involved in the learning process; relevant teaching methods and pedagogical techniques; personalized learning resources (both as a content and as a presentation); and suitable, learners friendly, motivating learning environment. Educational standard requires a logical and sequential organization of teaching/learning process. The outcomes are maximized when there is a precise structure of learning activities. In order to achieve relevant education for all students we consider as important the following requirements to the system supporting personalized education:

- providing a technology-based environment for successful teaching;
- enabling authorities to monitor learning process and outcomes;
- providing opportunity for the teachers, tutors and psychologists to determine the individual characteristics and preferences of the learner (personal student's model);
- enabling the teachers to build and structure the learning resources in various ways and create personal learning path consistent with the individual characteristics in the student's model;
- providing various ICT tools to enable a free exchange of views, experiences, information, resources, ideas among students and teachers.

The contemporary ICT offers rich opportunities for realization of functionality and management of e-learning system, of representation the learning resources in different and innovative way [7]. They supply adaptation techniques to meet the specific characteristics of each student and to exploit his / her strengths (imagination, creativity) to be trained to overcome weak points as concentration, organizing etc.

ICT-based Tools for Cooperation

The effective collaboration tools based on ICT are both social and content focused. They take advantage of technologies such as: participation (audio/ video conferences, blogging, online communities and networks); community police (defined rules, permissions and rights); easy spread of information and resources (profiles, tags, metadata structures); interaction among teachers and learners. The new technologies enforce considering not only content, but they involve also how that content can embrace various teaching and pedagogical methods, instructional and learning activities in the educational process. These tools:

- Ensure different forms of collaboration between teachers and/ or students:
  - Tools for communication – audio and video conferencing; idea, data and screen sharing; instant messaging and presence pointer; remote access, etc.;
  - Tools for creation of virtual communities – shared workspace, thematic networks, portals, communities;
  - Tools for manipulating common resources – databases, content management tools, mashup services, semantic web, tagging, file-sharing;
  - Context management tools – event scheduling, mind mapping, etc.
- Provide virtual space where the collaboration is carried out through cooperative services and resources;
- Assist educators in creating and reusing learning resources, good practice, etc.

Technology-enhanced Authoring Tool

ICT-based Authoring Tool supports various functions for management and creation of learning resources: development, structuring, storing, sharing, reusing, adaptation and personalization. The individualization of learning resources covers the following aspects:

- knowledge – regards volume, learning style, prior knowledge and competencies;
- presentation – regards additional audio and video illustrations, animation, simulation, 3D-modeling;
- contextualization – regards language, custom, traditions;
- type of learning resource – regards using interactive tasks, tests, etc;
- visual representation – regards font, colour, size, spacing.

All this functionality can be well integrated in and managed in the developed technology-enhanced e-learning environment.

Conclusions

The developed model of technology-enhanced educational environment is adaptive to the learner's characteristics and cognitive profile. Its practical application facilitates teachers by enabling personalization of learning resources and tasks; sharing effective pedagogic practice. Students also benefit because technology-enhanced environment easily links physical to abstract objects by using phenomenological approach. In addition they can take advantage of self-paced learning and unsupervised repeating practice. The ICT-based approach provides opportunity for revision of actions and using intrinsic feedback.

The presented model of technology-enhanced e-learning system:

- Supports personalized education and enables collaboration between its users;
- Teachers benefit from partnership with peers:
  - receive psychological and methodological recommendations;
  - create personalized learning resources in order to address different learning needs and students' preferences;
  - implement various teaching strategies and provide the created learning resources to achieve specific learning goals;
  - exchange teaching experience and good practices (pedagogical methods and tools, learning activities and resources).
- Students interact directly with teachers giving valuable feed-back to them.
- Students' requirements and preferences are addressed to the full extend.

As an extra advantage of the developed technology-enhanced e-learning environment is the possibility to collect educational data without much additional efforts (tests, questionnaires etc.). The data concerning the learning process can be collected, analysed and reported to policy makers. They can be used also by students, teachers, and institutions in order to evaluate the outcomes, educational resources and activities, pedagogical methods and tools.

References

1. Aroyo, L., Dicheva, D.: The New Challenges for E-learning: the Educational Semantic Web. J. Educational Technology and Society, 4, vol. 7, 59-69 (2004)
2. Bruslovski, P., Peylo, Ch.: Adaptive and Intelligent Web-Based Educational Systems. International Journal of Artificial Intelligence in Education, 13, 156-169 (2003)
3. Koper, R., Oliver, B.: Representing the Learning Design of Units of Learning. J. Educational Technology and Society 7, vol. 3, 97-111 (2004)
4. Andreev, R., Terzieva, V., Troyanova, N.: An Intelligent E-learning Environment Supported by a Communication Agent. In: 20th Int. Conference on Systems for Automation of Engineering and Research (SAER), pp. 248-253, Varna, Bulgaria (2006)
5. Andreev R., Terzieva, V., Kademova-Katzarova, P.: An Approach to Development of Personalized E-learning Environment for Dyslexic Pupils' Acquisition of Reading Competence. In: ACM Int. Conf. Proceeding Series, vol. 433. ACM, New York (2009)
6. Stages of Personalized Learning Environments, http://www.personalizelearning.com
7. Brown, J. S.: How to Connect Technology and Content in the Service of Learning. The Chronicle of Higher Education, vol. 55, issue 8, A120 (2008)

Some Simulation Models of Process' Synchronization for Educational Purpose

Svetlana Vasileva[1], **Konstantin Kapinchev**[2]

[1] Shumen University „Bishop Konstantin Preslavski", College – Dobrich, Dobrotitsa 12,
Dobrich, 9302, Bulgaria
svetlanaeli@abv.bg
[2] Technical University of Varna, 1 Studentska Str.
Varna, 9010, Bulgaria
k.kapinchev@tu-varna.bg

Abstract. The advantages of the simulation modeling makes it appropriate for using it in the education of students from different specialties. This paper considers simulation models of algorithms for concurrent access in the area of Operating Systems. One of the classical systems for simulation modeling GPSS World gives program means and environment for developing an innovative approach with applied nature in the demonstration and visualization of different algorithms. The methods for inter-processor communication in the algorithms for mutually excluding access to a shared resource are the object of this paper. The algorithms of Dekker and Peterson provide solution to the mutual exclusion problems in the case when two consumers, or agents, share single resource. These algorithms use only shared memory to create synchronization, which makes them easy for implementation and suitable for theoretical education

Keywords: simulation models, algorithms for synchronization, processes, education, operating systems.

1 Introduction

Computer simulations of systems and processes have significant number of applications in the real world. This paper focuses on the simulation and modeling and how it can be used in university education. The simulation model has the ability to be used multiple times during a number of classes. The data, generated by the model, can be used for statistical purposes and later on discussed and analyzed. The main application area of the simulation modeling is the Queuing Systems research. Also, being based on laboratory experience, simulation modeling, especially GPSS, appears to be suitable for academic education. Subjects like Operating Systems, Parallel Programming, Data Bases and others are example for that. These subjects has one common characteristic, they include process synchronization and parallelism in their curriculum.

This paper presents two algorithms, Dekker's and Peterson's, implemented in the environment of the *GPSS World Student Version* software tool.

2 Presenting of the algorithms for synchronization of processes by means of the program language C++

Process synchronization has a central place in the Operating System subject with topics like critical section and race for shared resource being studied in details. This paper reviews two algorithms of Dekker and Peterson [1], [2] and other researches, which are used to implement mutual exclusion. They achieve this by using shared memory for storing flag values. Proper presentation of the foundation of these algorithms, as well as others alike included in the Operating System curriculum, is a basis for better understanding the sequence of events and the deadlock avoiding technics. A number of different mutual exclusion algorithms can be found on the Internet [3]. In order to ease the learning process, these algorithms are implemented using well-known programming languages.

2.1 Dekker's algorithm

We consider a fragment from the C++ program which demonstrates Dekker's algorithm.

Example of a Computer Program from http://www.bg-informatika.com

```
shared int wants0 = 0; //this flag shows that P0 is to
//enter its critical section
shared int wants1 = 0; // this flag shows that P1 is to
// enter its critical section
shared int turn = 0; //gives the number of the process
//that is in turn to execute its critical section P0 :
while (1)
{ wants0 = 1;
    while (wants1)
      if (turn == 1) {
        wants0 = 0;
        while (turn == 1);
        wants0 = 1;
   }
  critical_section0();
  turn = 1;
  wants0 = 0;
  non_critical_section0();
}
P1 : while (1)
{ wants1 = 1;
    while (wants0)
    if (turn == 0) {
      wants1 = 0;
      while (turn == 0);
      wants1 = 1;
   }
```

```
 critical_section1();
 turn = 0;
 wants1 = 0;
 non_cricial_section1();
}
```

If two processes try to enter their critical section simultaneously, the algorithm will allow choose only one of them to do so. The choice will be based on the value of the common flag *turn*. The other will perform active waiting until the *turn* flag changes.

The wait loop is based on the flags *want1* and *want2*. They indicate the need of the corresponding process to enter its critical section.

2.2 Peterson's algorithm

A fragment from the C++ program is shown below in order to demonstrate the Peterson's algorithm.

Example of a Computer Program from http://www.bg-informatika.com

```
shared int turn = 0; //gives the number of the process
// that is in turn to execute its critical section
shared int interested0 = 0; //this flag shows that P0
//is to enter its critical section
shared int interested1 = 0; //this flag shows that P1
//is to enter its critical section
P0 : while (1)
{ interested0 = 1;
turn = 0;
while (turn == 0 && interested1);
critical_section0();
interested0 = 0;
non_critical_section0();
}
P1 : while (1)
{ interested1 = 1;
turn = 1;
while (turn == 1 && interested0);
critical_section1();
interested1 = 0;
non_critical_section1();

}
```

Most examples (internet, Computer Science books) are C/C++ programs, which create two processes. They have critical section, which is that part of their code where they access common resource. Each process remains in its critical section if it does not complete its instructions there, regardless of its current state. As a consequence of this the algorithms must ensure that the following conditions are met:

- Only one process can be in its critical section at one time
- Process cannot remain indefinitely in its critical section
- No process outside its critical section can prevent other one from accessing its critical section
- The decision cannot be made on assumption based on estimated speed of process

These conditions cannot be demonstrated by well-known languages while the operating system manages the processes.

3 GPSS models of algorithms for synchronization of processes

The models described in the previous point are static in comparison with the models that can be developed by means of simulation modeling. Simulation modeling allows studying of queuing systems (QS) with different types of input data and intensities of input requests and determining their main characteristics. The GPSS is known as one of the most popular programming language for modeling complex discrete systems [4] and [5]. It can be successfully used for modeling systems formalized as QS.

The objects of the language are the same as in the standard QS (requests, tools, queues and others). Sufficient number of such components allows constructing complex simulation models, keeping the primary QS terminology. This allows the usage of GPSS environment in a study process. It can be used by students to develop their own models and systems (computer systems, communication systems, Grid systems [6] and others).

3.1 GPSS model of Dekker's algorithm

Figure 1 shows the GPSS model of Dekker's algorithm. The variables *WANT1* and *WANT2* are cells with values, which will be checked by the transaction, the modeling process, for values equal to 0 or 1. If two processes are trying to enter a critical section concurrently, the algorithm will allow access only to the one for which the value of the variable turn is given.

If one of the processes is already in a critical section, the other one will wait until the first one comes out.

The cycle of waiting is realized by the help of two flags *WANT1* and *WANT2* which indicate the intention for entering the critical section and the variable turn that shows which one of the two processes must come in.

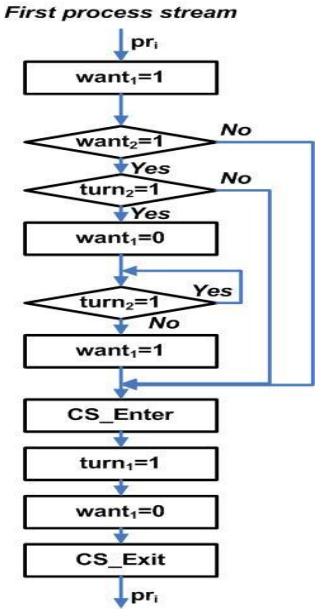


Fig. 3 Scheme of processing a process according to Dekker's algorithm: $want_1$ – flag that indicates that process 1 must enter its critical section; $want_2$ – flag that indicates that process 2 must enter its critical section; block *CS_Enter* – process pr_i enters its critical section; block *CS_Exit* - process pr_i comes out of its critical section; pr_i – process the number of which is i in the stream of processes.

3.2 GPSS model of Peterson's algorithm

The following is a fragment of the Peterson's modeling algorithm, developed in the GPSS World environment. This model corresponds with the one presented as *A2* algorithm [7] and is a GPSS „translation" of the algorithm from 2.2 concerning the stream of processes modeled with two streams of *GPSS* transactions.

Example of a Computer Program from Vasileva S. (2012) GPSS World model of Peterson's algorithm for synchronization

```
INTERES1 VARIABLE 0  ;Initial value of the variable WANT1=0
INTERES2 VARIABLE 0 ;Initial value of the variable WANT2=0
******************
TURN MATRIX ,2,1 ;the matrix is equivalent of the variable TURN
OBSTR MATRIX ,2,1  ;the matrix OBSTR models the shared resource
INITIAL MX$TURN(1,1),0
INITIAL MX$TURN(2,1),0
INITIAL MX$OBSTR(1,1),0
INITIAL MX$OBSTR(2,1),0
******************
GENERATE 5 ;generation of the transaction-process from the stream 1
ASSIGN 1,1  ;value 1 is given to the parameter 1 of the
            ;transactions from the first stream
ADVANCE 2 ;modeling of the initial processing of the processes
```

```
TRANSFER ,ObResurs1  ;to the part that models the Peterson's
                ;algorithm
GENERATE 5  ;generating of a transaction-process of the stream 2
ASSIGN 1,2     ;value  2 is given to the parameter P1 of the
            ;transactions from the second stream
ADVANCE 2   ;modeling of the initial processing of the processes
          ;right after entering
TRANSFER ,ObResurs2   ;to the part that models the Peterson's
                ;algorithm for the second stream
****************
ObResurs1   SAVEVALUE INTERES1,1  ;the process from stream 1
                        ;puts the flag INTERES1
MSAVEVALUE TURN,1,1,0  ;TURN for the processes from stream I is
    ;given value 0
Chaka1  TEST E MX$TURN(2,1),0,Vliza1   ;If TURN for the process
                                        ; from stream II is 1
   TEST E V$INTERES2,1,Vliza1   ;If the process from stream 2
                                ;puts its flag
   ADVANCE 1      ;the transaction will wait and
   TRANSFER ,Chaka1    ;try again to put the INTERES1 flag
Vliza1 SEIZE RESOU     ;The process enters its critical section
   ADVANCE 5  ;retention for processing in the critical section
   MSAVEVALUE OBSTR+,1,1,1   ;Record in shared resource of the
                                        ;process in stream I
   SAVEVALUE INTERES1,0  ;The process I takes the flag INTERES1
   RELEASE RESOU    ;then it goes out of critical section
  MSAVEVALUE TURN,1,1,1     ;and the process puts TURN
NAPUS1 TRANSFER ,NAPUS;The transaction process leaves the system
*******************
ObResurs2   SAVEVALUE INTERES2,1  ;the process from stream 2
                        ;puts the flag INTERES2
  MSAVEVALUE TURN,2,1,0  ;TURN for the processes from stream II is
                        ;given a value 0
Chaka2   TEST E MX$TURN(1,1),0,Vliza2  ;If TURN for the process
                                                ; from stream I is 1
   TEST E V$INTERES1,1,Vliza2   ;If the process from stream 1
                                        ;puts its flag
   ADVANCE 1; the transaction will wait and
   TRANSFER ,Chaka2   ;try again to put the INTERES1 flag
Vliza2  SEIZE RESOU   ;The process enters its critical section
    ADVANCE 5  ;retention for processing in the critical section
    MSAVEVALUE OBSTR+,2,1,1  ;Record in shared resource of the
                                        ;process in stream II
   SAVEVALUE INTERES2,0 ;The process II takes the flag INTERES2
    RELEASE RESOU    ;then it goes out of critical section
    MSAVEVALUE TURN,2,1,1    ;and the process puts TURN
*******************
NAPUS   TERMINATE 0  ;The transaction process leaves the system
    GENERATE 5000   ;Observation interval – 5000 model units
    TERMINATE 1
```

4 Visualization of algorithms for process's synchronization GPSS World environment

The modeling algorithms, described in previous point could be demonstrated easily in the corresponding windows of the *GPSS World* environment. These windows are possible to be requested to be visualized by the command Simulation Window from the *Window* menu of the *GPSS World* editor. These are the windows:

• *Blocks Window* - allows the chronological tracing of the transactions in blocks in model time. This window is shown in fig. 2;

• *Facilities Window* – shows: the number of entries in the corresponding device; The average stay time of the transaction in the device; Number of transactions waiting to take a place in the device and others;

• *Matrix Window* — shows results in values of the shared for the transactions resource in model time;

• *Savevalues Window* – window for the stored values. The change of the values of the flags *WANT1* and *WANT2* during the simulation of the Dekker's algorithm and the flags *INTERES1* and *INTERES2* during the simulation of the Peterson's algorithm is shown by means of this.

The rest of the windows that are possible to be called by the command *Simulation Window* will not be considered here because they are not used for demonstration of the algorithms of Dekker and Peterson.

Loc	Block Type	Current Count	Entry Count	Retry Chain	Line Number	Include-file
5 GEN	GENERATE	0	129	0	17	0
6 ASN	ASSIGN	0	129	0	18	0
7 ADV	ADVANCE	0	129	0	19	0
8 TRA	TRANSFER	0	129	0	20	0
OBRE...	SAVEVALUE	0	129	0	22	0
10 TES	TEST	0	129	0	23	0
11 TES	TEST	0	0	0	24	0
12 SAV	SAVEVALUE	0	0	0	25	0
CHAKA1	TEST	0	0	0	26	0
14 ADV	ADVANCE	0	0	0	27	0
15 TRA	TRANSFER	0	0	0	28	0
ZA1	SAVEVALUE	0	0	0	29	0
VLIZA1	MSAVEVAL...	64	129	0	30	0
18 SEI	SEIZE	0	65	0	31	0
19 MSA	MSAVEVAL...	0	65	0	32	0
20 ADV	ADVANCE	0	64	0	33	0
21 SAV	SAVEVALUE	0	64	0	34	0
22 REL	RELEASE	0	64	0	35	0
23 MSA	MSAVEVAL...	0	64	0	36	0
NAPU...	TRANSFER	0	64	0	37	0
OBRE...	SAVEVALUE	0	129	0	39	0
26 TES	TEST	0	129	0	40	0
27 TES	TEST	0	0	0	41	0
28 SAV	SAVEVALUE	0	0	0	42	0
CHAKA2	TEST	0	0	0	43	0
30 ADV	ADVANCE	0	0	0	44	0
31 TRA	TRANSFER	0	0	0	45	0
ZA2	SAVEVALUE	0	0	0	46	0
VLIZA2	MSAVEVAL...	65	129	0	47	0
34 SEI	SEIZE	0	64	0	48	0

Fig. 4. Fragment of the window of the GPSS World blocks during the modeling of Decker's algorithm.

After the model is prepared for execution with the command *Create Simulation* from the menu *Command* of the GPSS World editor, with the command of *Simulation Window* it is selected in which windows to observe the simulation of the model. After the command

START from the menu of *Command* is given, it is possible to see in the windows that are ordered on the screen the change in the values that we are interested in and the movement of the transactions modeling the processes in the system.

Fig. 3 shows such a combined „window" in which the execution of the model of Dekker's algorithm is seen. In the left „panel" it is possible to see the movement of the transactions-processes. On the top right is seen the use of the shared resource modelled by the *OBSTR* matrix and below it is the window of the *TURN* matrix where one can see the transactions from either stream put values either 0 or 1. On the bottom right corner is seen the rise/ subversion of the *WANT1* and *WANT2* flags. Above this window is the window of the single channel device that models the critical section for the processes from the two streams.

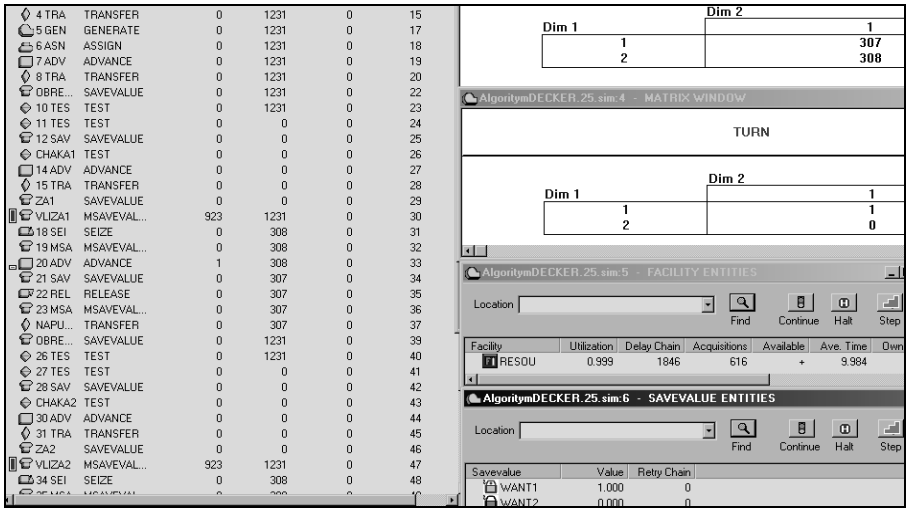


Fig. 3. Combined view to the windows for demonstration and tracing of the processes in Dekker's algorithm modeling.

Fig. 4 shows such a combined window for demonstration of the execution of the Peterson's algorithm. In this window one can: the change of the values of the *INTERES1* and *INTERES2* flags (right bottom); the change in the values of elements of the *TURN* matrix which models the turn variable for the processes of the two streams; the window for the single channel device that models a critical section (upper right); the window of the shared resource and the actions of change in it from the two streams (under the window of the facility *RESOU*). In a particular execution the values written by the first stream must be different with no more than 1 from the values that are written in the processes by the second stream.

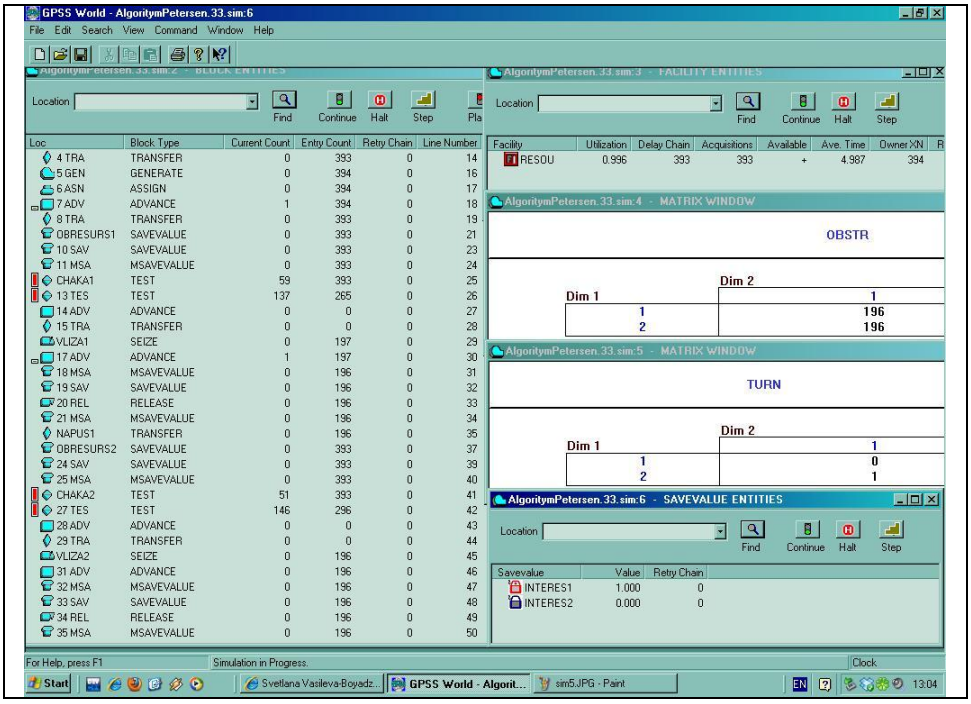


Fig. 4. Combined view to the windows for demonstration and tracing of the processes in Peterson's algorithm modeling.

5 Conclusions and future work

The models demonstrated and developed in the imitation GPSS World modeling environment present the opportunities for simulation of algorithms for educational purpose for mutually excluding access of processes to a common resource. The topic of synchronization of processes is an important issue that is considered in the discipline of Operating systems. Therefore, it is necessary the education in disciplines like this to be more attractive and updated with different demonstrations of the material being studied. After the necessary verification and adjustment the suggested models can serve for visualization of an inter-processor communication in studying other disciplines. In our opinion the GPSS World environment can be used for investigations in future for a broad class of algorithms for synchronization.

Thus, by means of a more detailed use of the opportunities for demonstration of algorithms of the environment for simulation modeling of GPSS World and their integration in the educational process, it is possible that the quality of teaching students might increase.

References

1. Silberschatz, G.: Operating System Concepts. (2006)
2. Tanenbaum, A., Woodhull, A.: Operating Systems Design and Implementation (3rd edition), (2007)
3. http://www.bg-informatika.com/2011/07/22.html
4. Modelirovanie system and setei svyazi na GPSS. http://en.radioland.net.ua/contentid-160-page1.html (2008), Моделирование систем и сетей связи на GPSS
5. Altaev, A.: Imitacionnoe modelirovanie na yazyike GPSS. http://www.ict.edu.ru, (2002) Алтаев, А.: Имитационное моделирование на языке GPSS.
6. http://ci.fmi.uni-sofia.bg/VasilGeorgiev/datagrid/grids13.pdf
7. Vasileva, S., Kapinchev, K.: Modelling of algorithms for mutual exclusion access to shared resources. (2011) Моделиране на алгоритми за взаимно изключващ достъп на процеси до общ ресурс. In. ICAICTSEE-2011, 2 -3 December 2011, UNWE, Sofia, pp.273-280.

Assimilation of geoinformational technologies in universities – perspective for future career

Galina Bezinska

Abstract. Geoinformational training programs are increasingly coming into higher education. In this regard, our article aims to present the current situation of GIS education at "Southwest University" Blagoevgrad. Some applications of these skills in other subjects are also mentioned. Using ArcGIS 9.3 software our students learn how to handle and analyze different spatial and statistical data and to prepare various thematic maps. In the further process the students successfully apply their knowledge and skills in developing research projects and not at least in developing their diploma thesis.

Keywords: higher education, teaching, GIS.

1 Introduction

At the end of the 20th century computer technology enabled the creation and application of spatial information systems for analyzing and generating solutions. The use of the largest communication system on Earth – Internet significantly increased. "All of these developments have created the conditions for the design, development and implementation of technology for storage, organization, management and use of spatial geoinformation with PC called Geographic Information System or Geoinformational Systems (GIS)" (Valchinov V. G. 2009). Today, practically there isn't any area of science where such technologies are not used.

Nowadays, GIS are the most common means of managing the huge and dynamically changing data streams (Kastreva, 2011). Their main purpose is to facilitate the user to expand its database and to manage the vast amount of information. Therefore the strategic efforts in education aimed at assimilation of the new technologies for creating, processing, presenting and analyzing of the data. This is essential for well trained staff to make important management decisions in every area of government.

Geoinformational technologies offer new and effective approach to solve a wide range of tasks in the socio-economics, geography, ecology, as well as in the management, analysis and forecasting of various critical events (fires, floods, earthquakes, etc.). The growing demand for high quality, reliable and convenient GIS engender respectively search of different specialists in the area of geoinformational technologies, from developer of software to highly skilled users. That's why, in the recent years, GIS take active part at learning process for most universities in Bulgaria.

The task of higher education in modern society is to prepare specialists knowledgeable, thoughtful, and skillful in modern informational technology, know how to acquire and apply independently knowledge in practice. In this regard it should be looking for new resources and opportunities for improving the effectiveness of the learning process. One solution is the introduction of. geoinformational training programs in universities such

as geoinformational mapping and GIS. Geoinformational mapping is automated creation and use of maps based on GIS and cartographic database. In fact it is an information-cartographic modeling of geosystems. (Berlyant, 1997)

2 GIS Training at Southwest University "N. Rilski"

In SWU „N. Rilski " the subject "GIS" was introduced in academic 1996/1997 year, but a significant step forward was made by changing the curriculum in 2004/2005 academic year. The same year has been equipped laboratory on cartography and GIS and has been purchased licensed software of ESRI. At bachelor course this subject is studied by the following specialties: "Geography", "Ecology and Environmental Protection " and "Pedagogy of geography and history". Unfortunately, this subject has the status "elective", despite of the growing importance and application of GIS in all professional fields. Generally, the number of academic hour of the subjects which are elective is reduced, especially for practical exercises. Thus it is difficult to place the students in a real environment and to cope alone with the tasks within 15 academic hours. For the future it would be appropriate to provide more practical exercises for this specialty. The exercises should be aimed at developing a new, closer to reality models of the environment. According Berlyant "Probably, not far the time when mapping holograms will create an illusion of reality and computer modeling will result in any difference between the map and picturesque scene." (Berlyant A.M., 1999)

As we mentioned above, lab classes are conducted in a specialized laboratory on GIS based on an interactive mode. Each student prepares several projects independently. Individual exercises entirely are related to the practical application of new knowledge, each of which is devoted to new topics. We should note that for some students this rhythm of work has been quite rapid, and they failed to consolidate their knowledge. In this regard, it would be better to provide free access to the laboratory in addition to usual consultations with teacher.

Basic theoretical knowledge of GIS is the subject of study in bachelor programs. Master's programs build on this knowledge and focus on the application of GIS to the specific specialization - in regional development, environmental protection and conservation of natural and historical heritage.

3 Applications of GIS knowledge

The lectures provide sufficient knowledge of how to perform analyzes in most GIS software. We must note that the students successfully apply these skills and improve analysis of spatial and statistical information, especially in the elaboration of their graduation diploma thesis. It should be given a special attention in a few student works. In addition to good cartographic design as a result of excellent knowledge and application of drawing tools, these works are characterized by interesting interpretations and conclusions made as a result of geographic analysis.

Among the simplest possibilities of GIS for analysis of data as selecting, listing and positioning of objects on the map, or measurement of lengths and areas, the use of different

methods for creating, transforming and modeling of statistical surfaces also are well absorbed. Students successfully apply GIS functions for various interpolation and transformation procedures, process and performance of regression and correlation analysis of statistical data to creating forecast maps.

GIS is used for various applications in different natural and social trends - geomorphology, climatology, hydrology, demography, etc. In the following lines we mention some examples with very good results at spatial and statistical analysis from different student works and diploma theses:

• Creating a geographic database and analysis of the 17th census; mapping a series of demographic maps and analysis of the dynamics of demographic processes in Bulgaria as a result of the migratory movements of the population, low birth rate and high mortality; Modeling of statistical surface in 3D space (Mapping the dynamics of demographic processes in Bulgaria);

• Relief modeling and its derivatives - slope and aspect of the slopes for Ograzhden and Belasitsa Mountains; performing geomorphological and geological analyzes of topographic surfaces, ("Geomorphological characteristics of the Ograzhden Mountain" and "Ecological and geomorphological research of Belasitsa Mountain ") (see fig.1)

• Analyzing of air quality (pollution levels of the main pollutants in the regions and points in excess of the threshold values for harmful substances and comparative analysis of emissions of pollutants per year) in the country and mapping a series of thematic maps; ("Mapping air quality");

• Analyzing the distribution of noise and damage from high levels of noise ("Noise in the country and in Blagoevgrad") (see fig.2)

Some student diploma theses, which are independent publications of various cartographic products, deserve special attention. For example, one of them design and produce thematic atlases; other realized tourist maps in the form of brochures; third prepare reference books and guidebooks on different themes. These cartographic products are result from the creation of geographical databases with different information. Here are several of the most interesting theses:

• Atlas “Biogeographic zoning of the land“ contains a series of maps and detailed information in Bulgarian and English language for biogeographic regions in the world and in Bulgaria;” (see fig.3)

• Atlas “Protected areas and sites in Blagoevgrad district” contains maps and a brief description of the national parks, reserves, protected areas and landmarks within the Blagoevgrad district;

• Atlas "Monasteries in Bulgaria" contains a series of thematic maps, where monasteries are classified in different categories. The purpose of presenting this type of information is to facilitate user of the atlas; (see fig. 4)

• Reference book "European Union" contains thematic maps on various themes with which the EU is presented in geographical, historical, demographic, economic, cultural and cartographic terms. It contains also geographically passport of EU member states;

• Guidebook "24 wonders of Bulgaria" has touristic character. It contains maps that provide interesting cultural, architectural and historical sites in Bulgaria in geographic and cartographic terms. The guide contains brief references of each of the twenty-four wonders of Bulgaria.

We must note also that a small group of enthusiastic students prepare individual projects to participate in EU programs. These contributions show that our students are

competitive; they have acquired the necessary knowledge and skills to implement new information technologies.

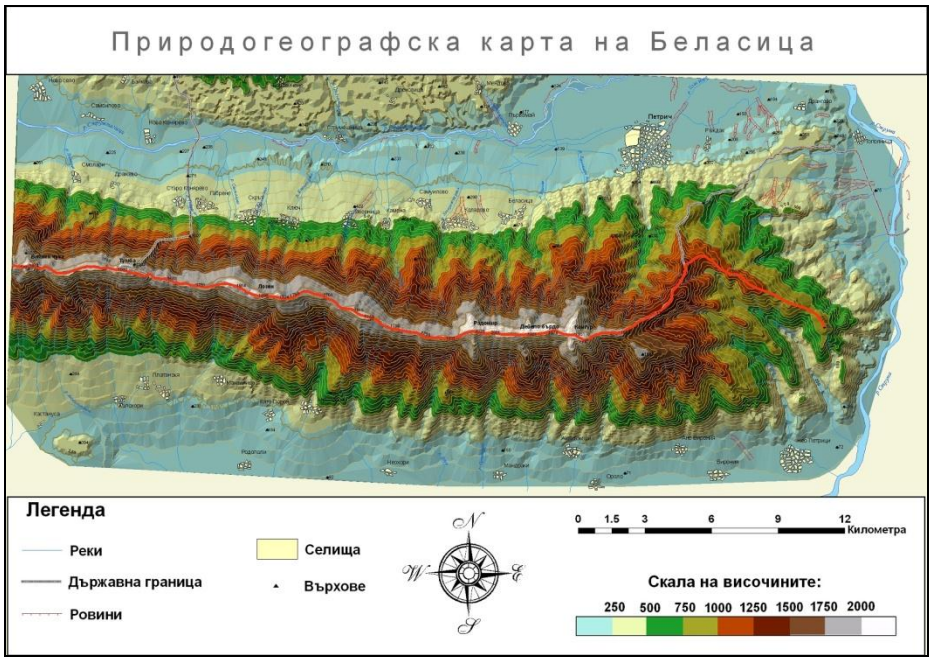


Fig.1. Natural geographic map of Belasitsa Mountain

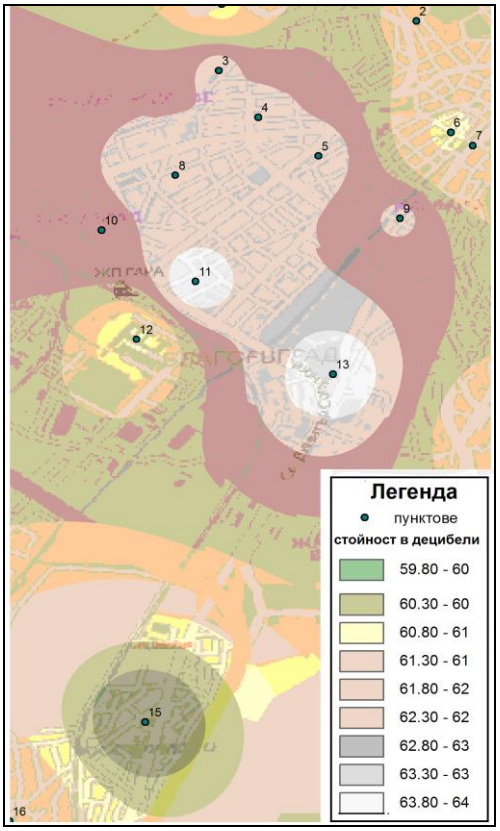
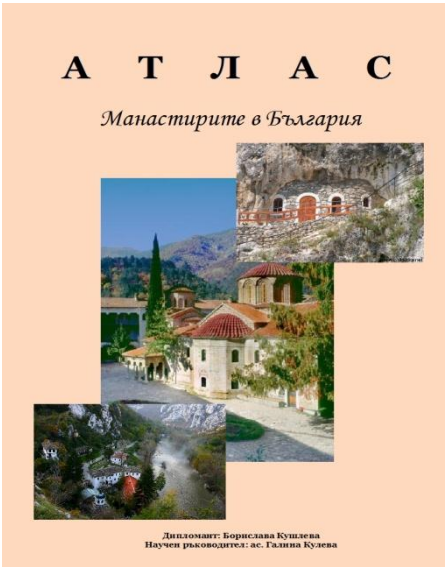


Fig. 2. Average noise levels in 2010 for Blagoevgrad

Fig. 3. Cover of Atlas "Biogeographical division of land "and one of the maps



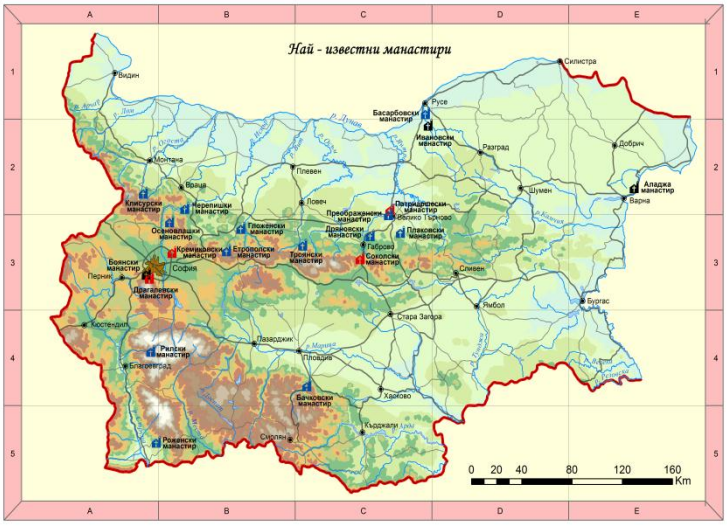


Fig. 4. Cover of atlas "Monasteries in Bulgaria and one of the maps in the atlas

3 Conclusion

In conclusion, it is clear that the contemporary development of geosciences is largely determined by the accelerated introduction of new methods and technologies using the latest achievements of science and technology. Naturally, the training of personnel must follow the process of technological development. Teaching GIS is only the first step in the utilization of

geoinformational technologies that are essential for future career of geographers, ecologists, pedagogues and others.

References

1. Valchinov V. G. Geoinformatics, UASG, second edition, 2009 – 296p.
2. Kastreva P. Geographic Information System and desktop cartography, University press, "N. Rilski", Blagoevgrad, 2011
3. Berlyant A.M., Geographic Information System in Earth sciences, Earth sciences, 1999
4. Berlyant A.M., Geoinformational cartography. M.:1997 –64p

Model for information security analyses in information security management system

Nedko Tagarev

Abstract. This article addresses the problems associated with permanent control after implementation of the management of information security governed by ISO. A model is proposed that incorporates process analysis, expert evaluation methods, analysis by scenarios and analyze the relationship between threats and safeguards. Certain are possible qualitative and quantitative changes in the protective mechanisms. There is an approach to normalize the data in order to improve the ability for analysis and reporting.

Keywords: model for analyses of IS, process analyses, defensive mechanism, treat, risk.

1 Reference model. Model Domenico Salvati [1]

In recent years, the application of laws and standards apply increasingly to provide information systems. Approach misleads companies to conduct expensive process control, excluding risk management. This gives the bureaucracy similar but controversial indicators of effectiveness and efficiency in terms of cyber threats. The management of the rule of law itself does not lead to control processes for information security. This can even lead to increased risk and uncertainty. Strict application of risk management promises effective and efficient selection of security mechanisms and counter threats to information security. [1]

Risk management is hampered by four problems:

1) Problem of uncertainty associated with the fact that interpretations of risk vary from community to community.

2) Problem of probability. In practice, the probability ratings are poorly accepted and subject to controversial discussions.

3) The problem of the influence of the working environment, which does not take into account the reasons for the likelihood.

4) Issue a decision. Security mechanisms defined by the selection criteria do not meet the rigorous approach of risk management.

The problem of uncertainty is solved at the state level by modeling scenarios of successive events. It is clear that threats and security mechanisms need to be treated as an indivisible whole. This leads to an understanding of the description of the threat from a certain probability distribution, and security mechanisms other. The problem is solved through curves.

The problem related to the influence is decided by pattern recognition. Of all the instruments are the most widely accepted "Rough Sets" methodology. It allows to draw conclusions based on small, incomplete and uncertain data. Decision problem is solved by adapting models of instrumental theory, which is not formally used in risk analysis. In most

cases, using best practices explored corporate experience. Based on this model consists of four modules, each of which is designed to solve one of the above problems:
1. Process module that reduces ambiguity in the description of the risk.
2. Function module introduces a common approach to assess the likelihood.
3. Modulus of the influence of context on the likelihood.
4. Module decisions determine preferences for a decision against the risks. It sets criteria for the development of security policies to counter threats. The four modules form a new and complete model for risk management. [1]

On sheme.1.is shown the entire model Salvati.

Technology for processing data [2]

Processing methods and knowledge that are directly related to security and defense decisions can be classified as:

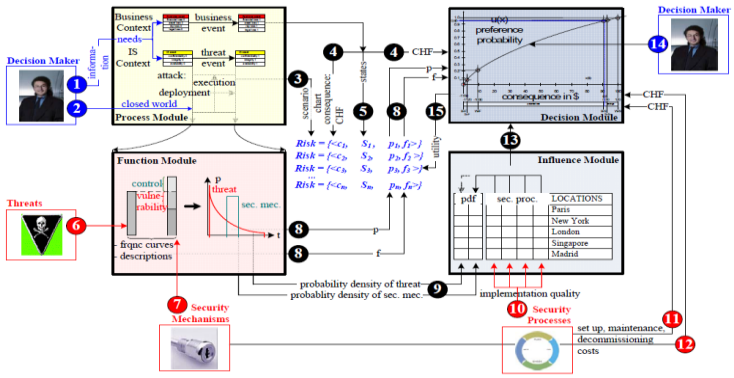


Fig. 1. Entire Salvati's model[1]

A. Qualitative methods:
• based on a subjective assessment (brainstorming, Delphi, historical analogy, benchmarking, panels, games, etc.).
• Structural (analysis cost - effectiveness, probability trees, morphological analysis, etc.)..
B. Quantitative Methods:
• interpolation, extrapolation;
• Probability (Markov models, Bayesian analysis, regression and correlation analysis).
• Operations Research (game theory, theory of stocks, linear programming, dynamic programming, non-linear and integer programming, network planning and management systems, queuing, etc..)
• Cause and effect analysis (econometric models, dynamic models of systems, etc.).
C. Mixed between A and B
D. Information methods:
• Classical (DB, man-machine interface, algorithms for processing messages for planning);
• Artificial intelligence (expert systems, natural language, pattern recognition and analysis of situations, self, etc.).
A trend towards applying methods A, B, C in the middle of the D.

2 Methods of analysis. Analysis by Scenario

Overall, scenario planning is a technique by which stimulate production of different perspectives and visions in a given area in order to make better predictions of development. Based on these ideas can be concerted and focused actions that lead to positive or negative impact on future development. Scenario planning differs from forecasting, in which is projected near future, based on extrapolation of past and present events. Scenario planning looks only to the future, but of course it has an impact in the past. Scenarios are influenced by the past, as each expert in developing their extrapolate, their past and current knowledge of actual development. [3]

Scenario analysis using a process of analyzing possible future events. It enables the construction of alternative versions of existing systems. The script does not attempt to show exact condition of the systems in the future. It presents some alternative future developments. This allows to monitor possible future results and conditions. Besides extreme conditions, this method allows you to monitor and analyze the roads leading to them.

Unlike forecast analysis methods using scenario analysis using the extrapolation of the past. It does not rely on historical data and it is not expected that the events of the past will affect the future yet. Instead, he tries to consider key developments and events that may be associated with the past.

Through the analysis of several scenarios indicate possible future results. This is useful for generating a combination of optimistic, pessimistic and most likely scenario. [4] Experience shows that it takes about three scenarios for further discussion and decision-making. Adding more scenarios make analysis more straightforward. [5]

An analysis by scenarios is intended to improve the process of decision-making. It also allows to take into account the results and consequences of decisions.

Process Analysis

The concept of "process" ... is given by the general theory of systems. A process as it determines the trajectory of a system in the field of its possible states. [6]

In economics and management processes that are related to human activity and are usually specifically designed to achieve a particular result ... process is a sequence of activities transforming inputs into desired output [6]. In complex process consisting of many actions, or those performed over a long period of time there may be several distinct intermediate result of changing inputs or change the system that runs research process. [6]

Experience shows that the use of a process approach leads to a number of beneficial effects, including:

• Reduction of production costs.
• Improving the efficiency of processes.
• Increasing the level of manageability, including improving the system of cost accounting, transparency in governance, accelerated procedures for decision making.
• Reducing the impact of human facto. [7]

3 Expert analysis methods

Brainstorming [8]

Brainstorming (MA), better known as brainstorming (English: brainstorming), is a group technique for generating new ideas beneficial decision making and enhance creativity. This technique can be used as an aid to: define the problem on which to work, identifying specific problems, finding solutions and sample testing whether they are reliable.

6-3-5 Method

Method 635 is one technique that is from creative techniques brain writing, which encourages a process of problem solving to produce new and unusual ideas in a group of people. It was developed in 1968 by a marketing and management consultant Bernd Rohrbach [9].

Investigation

The survey was writing consultation taking place generally in absentia, ie without direct and immediate contact between interviewer and respondents. It is relevant in two cases:

a / When you need survey the large number responsible for a relatively short time;

b / When respondents should carefully think over their answers given to a questionnaire printed eyes.

Surveys are useful for gathering large amounts of data. Less useful is the examination of a problem in depth

Surveys are rarely comprehensive (covering all members of the community being studied). More often they are selective. Therefore, the reliability of information received depends primarily on representativeness of the sample.

Model analysis ISMS

The proposed model in this paper for better analysis uses a combination of model for information security management, which is regulated by ISO 27010 (ISMS) and the model for the analysis of Domenico Salvati. This model is shown graphically in Figura.2.

The difference with Salvati model is that it is reported in much greater role of legislation, standards and prerequisites for information security. It should not be forgotten that the prerequisites are based on valuations that provide a new and unusual approach to solving certain problems. On the other hand, this model lacks the added complication of mathematical analysis methods.

The first step in this model described graphically in Fig. 2. is the introduction of the ISO standard. This process is described in detail in Ch. II. of this development.

The key point is to implement additional analysis model of the system, which includes an analysis of agents threatening to safeguards and impact of threats to assets. The location of the model in the Deming cycle is in function check. Using this model, perform the requirements of the ISMS for permanent control of the system.

Process Module

The aim of the process is to establish unit change in assets. Process analysis is a graphical representation of the chain of events. Displays the status of values by using techniques scenarios. These techniques allow a limited view of the future state of the organization's values. When it comes to information security for her values are confidentiality, integrity and availability. Events and counter an attack.

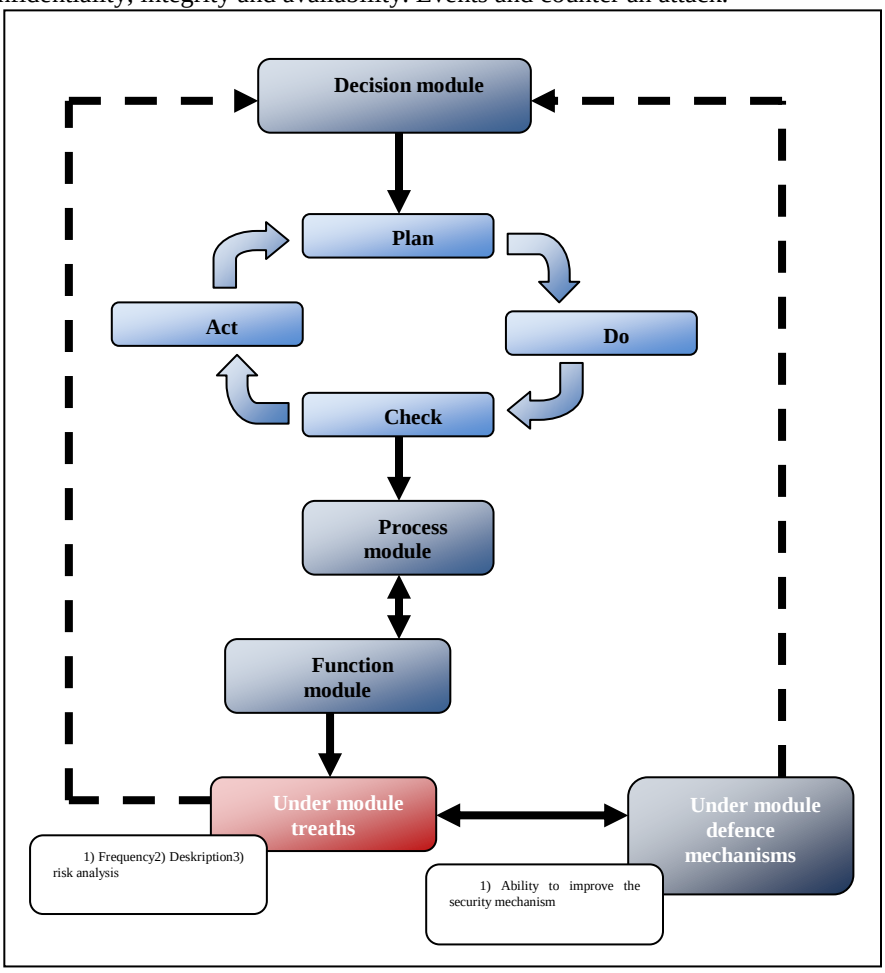


Fig. 2. General model for the analysis and improvement of protective mechanisms in ISMS

Process analysis allows to determine the impact and consequences in terms of values. The aim is that it be of assistance to decision makers. Additionally this support can be based on expert judgment. The process can be represented graphically. This is shown Sheme.3. When it comes to information security for her values are confidentiality, integrity and availability. Events and counter an attack.

Although one possible scenario analyzed are reported as other possible developments of the action. Scenarios are marked Sn, where S is the scenario, and n is the number of scenarios.

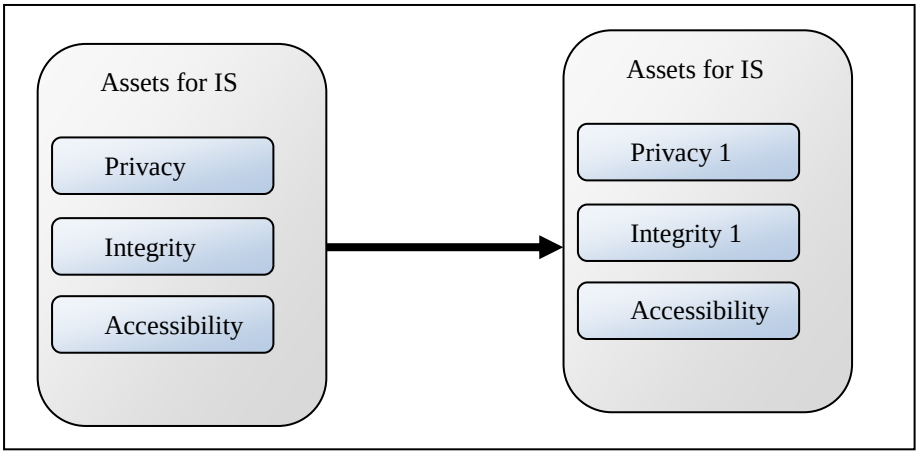


Fig. 3. Process for information security

The analysis will be carried out based on four main attributes which vary from attack described in the scenario. Each of the signs has three values.
1) Assets values of information security - confidentiality, integrity, availability.
2) Remedies - password, firewall (firewall), cryptography.
3) Economic values - finances, reputation, legal consequences.
4) Material values - tangible assets (a) casualties (b) and security (c).
(a) For assets means tangible assets - property, plant and equipment.
(b) Under casualties partially understood by victims to lose their lives. Viewing from the perspective of a terrorist attack aimed at inflicting damage, casualties and psychological impact.
(c) For security means psychological feeling of security. From security to fear.
To examine and assess the impact of each attack and the evolution of security and values using four indicators:
1) Probability of occurrence of the event - pk
2) Risk - rk. In cases where it can not calculate risk, suggesting that there is uncertainty. It is the worst option.
3) Effect of altered state of values – ik
4) Consequences of changing values - ck… k=(1..19)

Risk assessment of the threat

Risk is an economic concept. It is a natural manifestation of purposeful activity. Its implementation does not lead to the target and to deviations from expected results.

For this purpose, can be used two ways:
1. Komogoroff:
Risk = f[(pi),(ci)], where i is adverse event, (N) or
Risk =pi x ci (N)
2. Caplain и Garrik:
Risk = f[Sr,fr,pr,cr), where r = 1…n,
Sr -risk from scenario;

fr – frequency of the tread;
pr – probability threat to overpower protection;
cr – consequences.

Relationships between attack and defense system

To make this analysis should consider the strength of threat(t_i^s) and to be compared to the intensity of the defense mechanism ⟦DM⟧_i. For the purpose to analyze the knowledge of the threatening agent (k_i^a) and resources threatening agent (r_i^a). protection will be discussed as the set of properties to overcome the protective mechanism(R_i).

The aim of the proposed model is to increase the effectiveness of social protection in engineering or physical intrusion. Therefore, the most appropriate is to propose alternative remedies to be weighed against the force of the threat.

The knowledge and resources of threatening agents could previously only be inferred. Objectively to assess their level only achieved only attack regardless of its success or failure consequences. An analysis by scenarios is appropriate because it supposedly objective information about the knowledge and resources attacking agent. This is the case(t_i^s).

Another problem for the objective analysis can cause lack of a system to quantify the indicator knowledge. For this reason, you should use quality assessments - the agent has or does not have such knowledge. In mathematical representation of knowledge (normalization) when it has a value of 1,for YES and if not have a value 0. It's best that can be presented in tabular form (Table 1).

Table 1. Knowledge of the attacking agent

	Knowledge 1	Knowledge 2	…	Knowledge N N=1…
attacking agent	YES/NO	YES/NO	YES/NO	YES/NO

Contrary to knowledge resources can be quantified. Unfortunately, the amount of available resources is established. There can be found the resources that are used to carry out the attack. This can be done then, it has been realized. These resources can also be shown tabulated (Tabl.2.).

Table 2. Resources attacker agent needs to carry out the attack

	Resource1 Price	Resource2 Price	… Price	Resource M Price M=1…
attacking agent				

Safeguards can be measured quantitatively and qualitatively. Quantitatively measure the number of defense mechanisms and their qualitative properties. For a property that has a defense mechanism gets to, for property that has NOT received. Accordingly, in the

normalization of the data to obtain a value of 1when it's YES, and NO - 0. This can be presented in tables (Table.3.).

Table 3. Defense mechanisms and properties

	Property 1	Property 2	…	Property X X=1…
Defensive mechanism 1	YES/NO	YES/NO	YES/NO	YES/NO
Defensive mechanism 2	YES/NO	YES/NO	YES/NO	YES/NO
…	YES/NO	YES/NO	YES/NO	YES/NO
Defensive mechanism W W=1…	YES/NO	YES/NO	YES/NO	YES/NO

Interaction between attack and defense can be measured against the quality of attack power and capabilities of defense mechanisms. Is shown graphically in Graff.4.

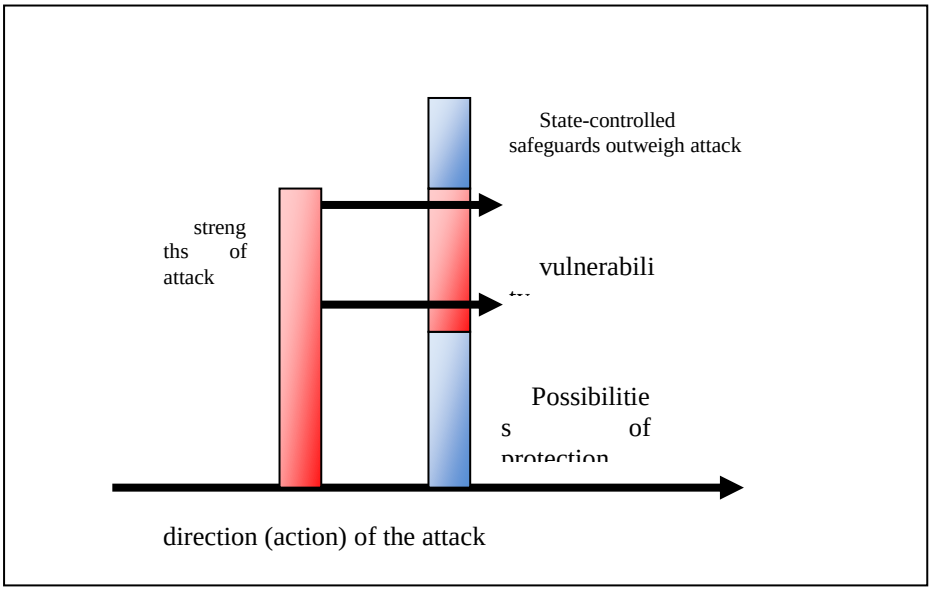


Fig. 4. Interaction between defenses and attack power

4 Conclusions

This article discussed a model of analysis can complement ISMS PDCA model, which is regulated by the ISO. From this model were taken for analysis methods that are the subject of other studies - mathematics and risk analysis.

Proposed an original model of analysis that can complement ISMS PDCA model, which is regulated by the ISO. Added analysis techniques to complement the ability of analyzing the capabilities and effectiveness of safeguards. Adding additional methods helps to verify assumptions about the possible threats, opportunities protective mechanisms and impact of threats.

To justify the model using the methods of process analysis to show the input and output change of assets, expert analysis methods - brainstorming, 6-3-5 and a survey to ensure the reality of the proposed model and analysis by scenarios which allows the analysis of the existing system.

References

1. Salvati,D. Management of Information System Risks, A dissertation submitted to ETH ZURICH for the degree of Doctor of Sciences, Zurich, 2008
2. Павлов, Г. „Информационни технологии в отбраната и сигурността", Университетско издателство „Стопанство", София,2003 г. стр. 95,96
3. Димитров,Д. „Приложение на сценарийното планиране в бизнеса, отбраната и сигурността". Издателски комплекс –УНСС, София, 2012 г, стр. 57.
4. Aaker, David A. (2001). Strategic Market Management. New York: John Wiley & Sons. стр. 108
5. Bea, F.X., Haas, J. (2005). Strategisches Management. Stuttgart: Lucius & Lucius. стр. 279
6. Пенчев, Г., „Развитие на въоръженията", Авангард прима, София, 2010г., стр. 55.
7. Цветков., Ц, „Методология на управлението на процеси при защита на критичната инфраструктура", в сборник "Защита на критичната инфраструктура в ЕС и България – икономически и организационни аспекти", Университетско издателство „Стопанство", София, 2010г.
8. Sutton, R, BusinessWeek, 26 July 2006
9. Rohrbach, Bernd: Kreativ nach Regeln – Methode 635, eine neue Technik zum Lösen von Problemen. Absatzwirtschaft 12 (1969) 73-

IT-innovations and monetary policy

Venelina G. Trifonova

Abstract. The key objective of the paper is to analyze the relationship between the IT-innovations and the monetary policy. With this aim the study put emphasis first on the issue about the money evolution under conditions of the IT-revolution, and second, on the monetary policy and the IT-innovations. IT-revolution encourages the most developed industrial countries toward the acceptance of new types of money and assets.

1 Introduction

In 1990s there is a sharp feeling about the economic influence of the new information technologies and on this base begins the theoretization for the impact of so called IT-revolution on the macroeconomic policy. The discussion about the effects of the financial markets' electronization put a lot of general theoretic questions, independently of the actual impact of IT-innovations on the economy. Methodological question is this about the nature of influence of the IT-progress in the field of communications and information on the monetary and financial sectors in developed countries. In the latter there are considerable effects of the above mentioned revolution and on this base interpretation of the IT-innovations on the economic and particularly monetary policy. The discussion on this important topic in the world literature put some new methodological problems as nucleus of the Post-Keynesian economics. In this relation however the studying of economic literature makes heavily the determination of what degree Post-Keynesian economists had been engaged into discussions about the influence of IT-innovations as regards the evolution of monetary theory and practice.

2 Money evolution under conditions of the IT-revolution

The discussion for the influence pf the IT-innovations on the economic policy is realized mainly in two different aspects, namely:

- Does the development of the new means for payment and the new clearing systems can influence the capability of the central bank to run effective monetary policy?
- Do IT-innovations presupposes the increased volume and better information distribution, approaching the markets functioning closely to the perfect information with the fully informed agents and eliminated informational asymmetry?

As concern the first aspect the appearances of new pay devices and clearing systems quite weakly conditions the effective monetary policy nevertheless that as a principle the large IT-innovations distribution leads to replacement of the traditional understanding of money with the trend to makes weak the regulative power of the central bank. The latter can be realized if the financial innovation process is made out of the framework of its current

state and presupposes new means for payment. However they should be independent from the conventional money consideration and to be administered by an institution different from the central bank. The new financial instrument has to replace the traditional money understanding in their function as account unit in the economy.

The replacement of the traditional money inevitably is confronted with substantial difficulties that finally emerge from the nature of the money. The latter are "social relation" and the transfer from one type of money to another can be done at the availability of economic, social and institutional process. The relative transformation of the types of money can not be done through the elementary shift in the technology. In their quality as social construction the money is not spontaneous result of the market functioning. What the economic and the social system see as money may change with the shifts in the system within the time. The transformation of one type to another can not be done simply under the influence of IT-innovations or because the agents think of the money as more efficient instrument. The generalization is that social and economic institutions as well the complex market relation stay in contrary of the acceptance of the new money.

There is a strong critic relation as a whole to the idea that the better informational distribution for the economic processes can conditions perfectly competitive situation in which is eliminated the need of monetary policy. This particularity is linked with the fact that IT-innovations practically can not overthrow the typical for the financial market risk for the financial transactions and decisions. On the contrary under conditions of more or less risky economic reality and speculative markets the use of informational technologies that facilitate the faster translation pf the information may event gives birth to higher degree of market instability and fluctuation of the development. Due to these reasons monetary policy and in general economic policy may even appear as more important in the future.

One of the obvious IT-innovations effects is the considerable growth of the quantity and value of the electronic transactions mainly in the developed countries. With this process are connected mainly the next phenomena:

- the distribution of so called "electronic money" known as "e-money" as a new means for payment and understood in their quality as instruments for keeping the cash value in connection with the payment among the agents without using the banking system;

- the IT-innovations' positive impact on the development of new clearing systems that do not presupposes the hold of reserves of the commercial banks in the central bank.

The faster translation of economic information with lower expenses may make easy the creation of private clearing systems. The payments on settlement may be done by the banks through the use of what C.Freedman called "monetary net", i.e. reserved funds in some electronic form that can be transferred through the communication nets (Freedman, 2000: 211-227) The large distribution of these transformations of the payment system will lead to decline of demand of the traditional money and particularly of the monetary base in form of bank reserves in the central bank.

About the question could the IT-innovations eventually condition the central bank's impossibility to pass effective monetary policies there are to different scientific positions:

- progressive under which the financial innovations born from the electronic revolution may substantially to reduce the demand for traditional money without these innovations to undermine the role of the central bank.

- conservative under which the IT-innovations can reduce the central bank's capability to influence the economy through the monetary policy.

3 Monetary policy and IT-innovations

According to the mentioned progressive position of the link "IT-innovations – central bank's monetary policy" in the literature is said that a crucial factor for the influence of the payment system's transformations over the dynamic of the traditional money demand id the possibility for decrease of the bank reserves' demand in the central bank, i.e. the demand for monetary base. But even the great demand reduce is difficult to has as a consequence the big decline of the regulative central bank's power. M.Woodford wrote in connection with the interpretation of the situation for the run of the monetary policy into an imaginer world that lacks money that without the low degree of reserves demand the central banks through the monetary policy can impact the economic development. This they can do in a better way through their typical impact on the rate of interests (Woodford, 2000: 229-260). The central banks can realize the adequate activity trough the payment of interest on the commercial banks' reserves.

According to the conservative position the connection "IT-innovations – monetary policy of the central banks" is that the adequate banks in the future can lose their capability to intervene with the monetary policy. The reason is that the IT-innovations create shifts in the controllable variables of the central banks. This does not lead to changes in the relative variables for the national economy as whole. Nevertheless that the central banks' transactions on the open money market are very small part from the general financial transactions these banks still could influence the interest rates. This is due to the fact that their transactions are fundamentally different from all others, namely:

- the securities purchase always means increase of the accounts for the bank reserves of the seller;
- the securities sales always means decrease of the accounts for the bank reserves of the buyer.

Because of the specific role of market participant the central bank has a monopoly over the ensuring of reserves. Thus it can influence the interest rates without engaging with big financial transactions. It is well known for the market participants that when the central bank made up its mind to make this it could engage itself into grater transactions and by this way to fix the required level of the rates of interest.

In connection with the building of the market concept for the money J.Hicks wrote that the important meaning of the central banks' liability side of the balance sheet can be expressed through the determination of the fully real money, while the other means of payment are quasi-money (Hicks, 1989). In other words the financial instrument as account unit in the national economy at the same time is its own payment device. The account unit in the national economy is not simply a figure for calculation. There is a positive demand for it that might reach considerable levels. In the economic practice are known situations into which the agents are looking for fully real money instead of quasi-money. The similar is the case when there is a strong decline into degree of agents' confidence because of the heavy economic shocks. Under such situations the agents consider the liability side of the central bank's balance sheet as an instrument involving the greatest confidence.

As a principle the IT-revolution can stimulate predominantly the most developed industrial countries toward the acceptance of new types of money with the well known functions of exchange mean and account unit. After the new monetary instrument reach "critical" level of distribution the resulted net effects can presuppose their acceptance in the frame of total economy. In this relation there are several obstacles I front of the money

replacement. For example the overthrowing of the replacement of earlier existing (old) money although that they appear as less effective financial instrument in comparison with the new money. Approach of this kind may be due to higher costs born from the replacement of one type of assets with others. As a consequence of that the degree of competition when different monetary instruments are used is not very strong. But it should be weighted that the replacement of the central bank as privileged place for settlement of the non-balanced payments of the commercial banks will meet considerable difficulties due to next reasons:

a) The central banks are non-risky and can be lenders of last resort.

b) The private banks act as settlement agents for all others banks but this type of banks are risky and that is why the rest banks will not feel them rationally in aspect of its competitive advantages.

c) It is not possible for the central banks to fall in the position of passive observers of the spontaneous market revolution.

To reduce the risk arising from the different innovations in the financial system and which more or less endangers their regulative strength the central banks very strictly watch the financial markets and undertake different measures to control their evolution. These banks may go ahead toward the formation of new regulations with the help of the governments. Thus the replacement of traditional money and the eventual of reduce of the central banking meaning nevertheless as possible principally looks predominantly as far away options in the time.

The presented viewpoints can be expressed with more general notions, as follows:

- the money by their nature are "social relation" instead of simple technical device for making payments;

- the emergence of the money as account unit and mean of exchange is a logical result from the use of a complex of social and economic processes and is not simply product from technological shifts or from an optimization behavior of the individuals.

With similar thesis can be explained the central banks' existence and functions. The confidence is the major factor when is undertaken the clarifying of the money nature and their origin. Given financial instrument is accepted for money as much as the agents in the economy believe that this instrument will be accepted from any other individual. Historically universally accepted instrument used ultimately as a mean for payment became a liability of the central banks. Actually they can propose guarantee for the universal acceptance of the money as no other financial institution can do it.

The acceptance of new types of money and new institutional creation for their administration are economic, social and institutional processes. Historical analysis of the money and the central bank's monetary policy discloses the complex nature of the processes that brought to the contemporary monetary systems into the economically developed world.

References

1. Freedman, C., "Monetary Policy Implementation: Past, Present and Future – Will Electronic money Lead to the Eventual Demise of Central Banking", in International Finance, 2000, 3.
2. Hicks, J.R. 1989, A Market Theory of Money, Oxford: Clarendon Press.
3. Woodford, M. "Monetary Policy in a World without Money", in International Finance, 2000, 2.

Applying the information technologies in the financial and business analysis of small and medium size enterprises' business in Bulgaria – problems and perspectives

Rositsa Ivanova

University of National and World Economy, "Accounting and Analysis" Department
Email: Rosi_Ivanova@abv.bg

Abstract. In the conditions of globalization and market competition, the requirements in relation to the information quality and usefulness for making prompt and correct decisions in the process of enterprise management are significantly increasing. Being considered as a practical expression of specialized management function, the financial and business analysis provides important information to the managing body. Modern retrieval of efficient analytical information is an important prerequisite to make good management decisions and successful business development.

This publication highlights some problems when small and medium size enterprises in Bulgaria use information technologies. This issue is quite topical due to both the specific conditions and the legal regulations for the business in the country, and the perspectives for each enterprise's survival in the conditions of severe market competition and impact of the financial and economic crisis.

The aim is to define the problems relevant to the use of information technologies in the practice of financial and business analysis of small and medium size enterprises' business. On the basis of the above, the publication defines different options for analysis of the business of these enterprises (methodology and methods for marketing, innovation, investment, production, sales and financial operation analysis) when using the functions of modern information technologies.

Key words: information, analysis, problems, financial position, decisions

Introduction

Information technologies are world economy most rapidly developing sector. They are significantly applied in business, which is to a great extent dependent on information technologies. Due to the constant improvement of information technologies and the introduction of new high-tech applications, computer systems become more and more difficultly controllable and understandable. This fact also affects business.

The establishment of value for business and its further increase requires objective integration between different systems of interrelated indicators characterizing the financial stability prospects, market success and efficiency of business in the situation of globalization, sever competition and financial and economic crisis impact. The need to use integrated systems of financial and non-financial indicators for balanced analysis of business efficiency in different sectors: finance, customer service, internal business processes, human resources and human resources development, increases.

The system of key indicators for analysis of business efficiency plays a major role in the management of business subjects' operations. New ideas about the efficiency analysis and evaluation methodology are established at the end of XX century. On one hand, financial managers and analyzers consider the balanced analysis of business efficiency by using both financial and non-financial indicators. On the other hand, economists stand positions for wide use of enterprise's value management based approach. Thus at the end of XX and at the beginning of XXI century, in the conditions of knowledge based economy, two modern

approaches for business efficiency analysis are established. We believe that such approaches may be defined as management and economic (financial) approaches.

Management approach is based on using a balance system of financial (profit, profitability, liquidity, etc.) and non-financial (management quality, ability to achieve the selected strategy, etc.) indicators for efficiency analysis. This system should be integrated with the traditional efficiency analysis' indicators used by the financial business analysis. Such integration allows performing business strategic analysis and management, whereas the aim is to increase the efficiency of enterprise's operations and to realize the strategy it has selected to implement. Analysis is performed by means of a system of indicators and determining factors, and the resulting analytical information and conclusions made are grouped in different sectors of the balanced system of indicators. Its interpretation allows to define a system of key efficiency indicators, whose meanings and dynamics are used for measuring the level of attained specific objectives within the separate inter-balanced sectors of the system. Correctly selected key indicators are included in the strategic card and are joined by casual relationships objectively existing among them.

The economic approach (we think it's correct to define it a financial approach) is based on the use of different financial indicators, such as economic value added (EVA), market value added (MVA), shareholder value added (SVA), cash flow return on investment (CFROI), cash value added (CVA), etc. for the evaluation of business. The prospective in this area is to achieve a future integration with the balanced system of indicators for effectiveness analysis.

Under this new situation of business operations, there is an objective need to establish a common automated system that can simultaneously serve all information necessities of both financial department personnel and of the other units within the enterprise. Such integrated business efficiency analysis approach, based on the abilities of modern information technologies and on the methodological and technical characteristics of financial-business analysis, is a premise for making correct, timely and justified management decisions for business development in operating and strategic aspect.

Main text

Modern approaches for enterprise's operations and business effectiveness analysis impose the need to use modern information technologies in analytical and business practice in general.

Shortly describing the information technologies use in the financial and business analysis practice, we can come to the following conclusions:

a) At the end of 60s of the XX century, people began using computer equipment for automation of different aspects of the enterprise's operations. Different automated systems are developed and implemented in practice during this period. The system for material requirements planning (MRP) is one of these systems. The system operation is based both on the Bill of Materials (BOM) and on the Master Production Schedule (MPS) of the enterprise. Such integration is fully justified since the products' specification is an expression of the need for input components required for their production, and the production schedule (or the business plan) contains information about the volume, range structure, quality, complexity and competitive power of products that are subject to production and sales. The integration between the automated BOM and MPS systems provides information about the establishment of optimal specification of products with view of planning the requirement of

materials for the production of planned quantities in accordance with MPS. Thus, the information about the requirement for materials is essential both for arranging sales (management of relations with clients) and for maintaining stocks by types of materials (management of specific type of internal business process) and arrangement of supplies (management of relations with suppliers). The use of the MRP automated system during the last thirty years of the XX century was useful with view of: maintaining optimal amount of stock of materials; reducing the amount of unfinished goods; increasing the production effectiveness; attaining equality of production and sales. However, the system has a significant disadvantage. It is intended only for the automated monitoring of material resources requirement, without considering the relation with other production resources in the enterprise, and namely fixed assets and personnel.

b) In 70s of the XX century, the MRP's functional abilities are expanded through the development and implementation of capacity requirements planning (CRP) module in business. The relation between CRP and MPS allows to planning the enterprise's fixed tangible assets required for completing the business plan by volume, range structure, quality and competitive power. The integrated use of MRP and CRP puts the beginning of the use of automated systems for closed loop material requirements planning (Closed Loop MRP).

c) A new class of automated systems for manufacturing resource planning (MRPII) was developed and used in business during the 80 years of the last century. In contrast to MRP, the MRPII system is used for planning all resources of the enterprise. As a result of the information technologies development at the end of the XX century, the improvement of MRPII system and its functional expansion, as well as the increase of demand for a common automated system of the enterprise with view of its operations management, ERP (Enterprise Resource Planning) systems occur. They find successful and even expanding application in real practice for the purposes of achieving the goals and solving the tasks of the balanced business analysis and of establishment of efficient analytical information with quality characteristics that are required and sufficient for making correct management decisions.

ERP systems may be defined as information systems for identification, planning and management of enterprise's resources needed for the implementation of its operations, and the emphasis is put on the interrelations and dependences among its different elements (marketing, innovation, production, trade and finance) in the process of performance of commitments to clients and satisfying their requirements.

The main functions of ERP systems within the practice of financial-business analysis may be grouped in the following areas: a) Establishment of database for constructive, technical, technological specifications of products that are manufactured and offered by the enterprise. b) Planning the requirements of material resources by types, terms of delivery, suppliers and other factors in relation to the performance of the production and sales business plan. c) Capacity requirements planning (CRP), including the availability and use of fixed tangible (non-current) assets by types and in general for the enterprise in relation to the performance of the production and sales business plan. d) Planning the enterprise's personnel availability (number, structure, education, qualification and improvement, turnover, motivation, etc.) in relation to the performance of the production and sales business plan. e) Development of business plan for the amount of production and sales, with possibility to analyse its performance in relation to amount, rhythm, range structure, quality, complexity and competitive power of products. f) Establishment of database about the enterprise's expenses, the cost of products and the full cost of sold products, works, services. h) Establishment of database about enterprise's revenue classified on the basis of different

criteria. i) Establishment of common information database for analysis of financial result's amount and dynamics depending on the mechanism of its establishment. j) Analysis and management of business efficiency by using a system of key efficiency indicators. k) Establishment of information database for analysis of enterprise's financial position in the following areas. l) Establishment of information database for analysis and management of internal business processes with view of their optimization and improvement of their efficiency. m) Establishment of client database for the purpose of increasing the efficiency of processes that are directed to the customer relationship management (CRM).

d) At the beginning of the 21st century information technologies become more mobile. It is possible to use the so called "cloud technologies" for organization of financial and business analysis of the enterprise's work. To the opinion of the US National Institute of Standards and Technologies, cloud computing is a model for enabling convenient, on-demand network access to computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.[5, p.59] Cloud technologies are composed of three service models: SaaS (service as a service, whereas the enterprise pays to use a specific software application in the cloud), IaaS (infrastructure as a service, whereas the enterprise pays for the use of computing resources in the cloud, for example: process time, computer RAM, network computing, etc.) and PaaS (platforms as a service, whereas the enterprise rents both infrastructure and software applications in the cloud in order to make its own).

Four deployment models of cloud services are available: private, community, public and hybrid cloud. The private cloud infrastructure is owned or rented by an enterprise and is used by this enterprise only. The community cloud infrastructure is shared by several enterprises and is designed to support a specific community of users having common mission, strategy, organizational rules and information safety requirements. The public cloud infrastructure is owned by one enterprise which sells cloud services. The hybrid cloud is a combination of two or more clouds, which can be differentiated regardless of the fact they are inter-joined.

When organizing the financial and business analysis of the enterprise's business for input information processing, the capacities of the private cloud may be successfully deployed. It is accessible only within the frame of the enterprise. The information protection level, which is created and improved on continuous basis within the enterprise itself, is the main reason to establish a private cloud. However, the establishment of an IT department is quite expensive for small and medium-size enterprises. The impact of the world financial and economic crises, the adverse processes within the economy, the significant indebtedness of business, and some other factors prevent these enterprises from increasing their expenses to be incurred for acquisition of modern information technologies. Here comes the issue for using external cloud technologies for information processing in general, including the analytical information, which is so important for management decision making.

The external cloud computer services may be defined as outsourcing for storage and processing of information, which is considered a specific type of enterprise's fixed intangible asset. The distribution of cloud technologies solves a number of organizational and technological challenges the enterprise is facing. The main challenge is related to the information security.

ISO-27001 international standard defines information security as a continuous process of information confidentiality, entirety and accessibility safe keeping. The texts of ISO-27001 [4] and BS 25999-2 standards, which regulate information security, are quite

interesting here. Different requirements for information security are introduced in case of information processing and storage by external providers (to our opinion, cloud service providers). Most important among them are: a) maintaining the enterprise's information security system and information processing facilities in good working condition (control A6.2.); b) identification of risks to the enterprise's information and information processing facilities (control A6.2.1.); c) addressing information processing issues in agreements that cover the access, processing and management of provided information, the information processing facilities used, the used products and provided services, whereas describing all information protection methods and means in details (control A6.2.3.); d) information access management (control A7.2.1.); e) creation of information arrays backup and software provision (control A10.5.1.), etc.

Cloud Security Alliance experts identify seven categories of threats that might affect the information security. They are classified in the following groups [7, p.41]: a) abuse and nefarious use of cloud computing; b) insecure application programming interfaces; c) malicious insiders; d) shared technology vulnerabilities, i.e. it is not always possible to ensure reliable isolation of information arrays owned by different enterprises – customers of the cloud service provider; e) data loss or leakage due to errors of both the customer's and the cloud service provider's staff; f) account or service hijacking, computer viruses, computer equipment or software malicious damaging; g) unknown and unidentified risks, for example the fact that enterprises are not having full information about the cloud environment and the risks related thereto.

We believe that cloud technologies have a number of benefits that might be successfully used by the business for input information processing and for obtaining effective analytical information about the financial results, the financial position and its diagnostics, for the financial equilibrium and for the business efficiency in general. Most significant among them are: a) resource saving; b) maintaining of assignment priority within the cloud; c) improving accountability.

The studies show that business identifies the major factors affecting the risk occurrence when using cloud technologies. The impact of these factors is defined as follows [7, p.7]: a) unclear rules for defining the cloud service provider's security policy (32%); b) inadequate staff training and imperfect IT audit (19%); c) provider's website privileged access imperfect control (15%); d) closeness of information to the information of other enterprises within the cloud (11%); e) unclear options for information recovery if needed (9%), etc.

The risk of using cloud technologies for information, including analytical information, processing and storage, may be minimized by taking measures in different aspects: a) preliminary analysis of the cloud service provider's work (for example, productivity analysis, studying the curriculum vita of its owners and IT managers; introduced information security and business continuity policies and procedures; financial stability analysis; existing legal risks, etc.); b) providing for specific information security clauses in the contract with the provider emphasizing on problems identified during the preliminary risk analysis; c) safe keeping of information backup at the enterprise; d) development of strategy for the returning already processed (or already archived) information to the enterprise (re-in sourcing) in case of problems with the cloud service provider; e) ensuring alternative cloud service provider; f) implementing permanent supervision on the compliance of the contract's protection clauses and their implementation when processing and storing information from the provider.

Information technologies – analysis in Bulgaria

In Bulgaria, business does not completely use the options of the modern information technologies in the process of developing and making management decisions for enterprises' development. Data from the National Institute of Statistics about the use of information technologies in business are summarized in the tables below:

Table 1. Enterprises using internal electronic exchange of data [8]

Enterprises	2008		2009		2010	
	%	no	%	no	%	no
1. Total, including:	35.4	8781	33.9	8980	34.1	10389
a) 10 to 49 employees	33.0	6581	30.5	6276	31.0	7693
b) 50 to 249 employees	43.4	1836	43.8	2278	46.3	2279
c) more than 250 employees	54.0	364	56.8	427	60.5	417

Table 2. Enterprises whose business processes are automatically connected to those of suppliers and/or clients [9]

Enterprises	2008		2009		2010	
	%	no	%	no	%	no
1. Total, including:	14.4	3565	14.1	3752	18.1	5496
a) 10 to 49 employees	13.1	2614	12.2	2507	16.4	4073
b) 50 to 249 employees	19.0	806	20.4	1059	24.8	1219
c) more than 250 employees	21.5	145	24.8	186	29.5	203

Table 3. Enterprises using software applications for customer relationship management (CRM) [10]

Enterprises	2008		2009		2010	
	%	no	%	no	%	no
1. Total, including:	14.0	3466	12.9	3421	14.4	4387
a) 10 to 49 employees	13.0	2590	11.1	2277	12.8	3187
b) 50 to 249 employees	16.6	704	17.6	914	20.7	1017
c) more than 250 employees	25.5	172	30.6	230	26.6	183

Table 4. Enterprises using automated exchange of data with external information and communication technologies [11]

Enterprises	2008		2009		2010	
	%	no	%	no	%	no
1. Total, including:	41.5	10293	37.7	10003	38.1	11605
a) 10 to 49 employees	38.3	7626	33.7	6929	34.6	8588
b) 50 to 249 employees	53.9	2279	50.8	2643	53.0	2609
c) more than 250 employees	57.3	387	57.5	432	59.2	408

Table 5. Enterprises using ERP systems [12]

Enterprises	2008		2009		2010	
	%	no	%	no	%	no
1. Total, including:	5.7	1406	7.7	2051	10.8	3302
a) 10 to 49 employees	4.0	797	5.6	1152	8.4	2077
b) 50 to 249 employees	10.1	428	13.0	677	19.6	962
c) more than 250 employees	27.0	182	29.5	221	38.0	262

The relative number of small enterprises, with 10 to 49 employees, that use ERP systems is still quite low (8.4% during 2010), and in comparison to the previous two years we can see a minimum increase of this share. The ERP systems is more intensively applied by the medium-size enterprises, with 50 to 249 employees, in the country (19.6% in 2010), whereas for the large enterprises, with more than 250 employees, these systems are quickly introduced in business (38% in 2010).

The limited use of ERP systems in small and medium-size enterprises in the country is due to a number of objective reasons: 1) In the conditions of financial and economic crises, enterprises do not have sufficient financial resources to invest in information technologies. The impact of this objective fact is further increased by the high extent of indebtedness among enterprises in the country. 2) Omissions in the organization of making employees familiar with their duties in relation to the use of information technologies, the benefits from their use, and the information security of the enterprise. 3) Omissions or delay in the establishment of information security system in the enterprise. 4) Accidents occurred in relation to the information technologies security due to different reasons: hardware or software faults; external attacks; viruses; incompetent use; subjective errors; pharming, phishing, etc.

Conclusion

The application of modern information technologies in the practice of financial-business analysis brings a number of benefits for the enterprise and the efficiency of its business. They may be classified in the following areas: 1) Operations are optimized through

curtailing the unnecessary and routine operations thus reducing the expenses. 2) A single entry of initial information is achieved. 3) The performance of enterprise's employees is improved. 4) Information technologies have convenient interface, high working speed, quality and reliability, they are flexible and user-friendly. 5) The amount of expenses for introducing them in real business is affordable. 6) There is a good introduction expenses/ benefits of use ratio. 7) They provide reliable security of information in relation to different areas: physical, documentary, cryptographic, etc. 8) They give opportunity to make correct management strategic and operating-tactic decisions for the enterprise's development.

The use of modern information technologies in the practice of financial-business analysis, and in business in general, is an objective prerequisite for increasing the enterprise's effectiveness and competitive power. The ERP systems' capacities within the common information space allow the financial managers to obtain the authentic effective analytical information required for making correct and justified management decisions for business development.

References:

1. Грязнова Г., А.Федотова, "Оценка бизнеса", М., "Финансы и статистика", 2006
2. Каплан Р., Д. Нортон, "Сбалансированная система показателей: от стратегии к действиям", М., ЗАО "Олимп-Бизнес", 2003
3. Рассказов С., А.Рассказова, Стоимостные методы оценки эффективности менеджмента компании, Финансовый менеджмент №3/2002
4. Международный стандарт ИСО 27001. Information technology. Security techniques. Information security management systems. Requirements.
5. Georgiev G., Cloud Technology – Role and Significance in the Development of NII in the Ministry of Defense and the Bulgarian Army, CIO, 07.2012
6. Rappaport A., "Creating Shareholder Value. The New Standard for Business Performance", New York: Free Press, 1996
7. CIO, www.cio.bg
8. http://www.nsi.bg/ORPDOCS/ICT_Ent2.3.5.xls
9. http://www.nsi.bg/ORPDOCS/ICT_Ent2.3.6.xls
10. http://www.nsi.bg/ORPDOCS/ICT_Ent2.3.3.xls
11. http://www.nsi.bg/ORPDOCS/ICT_Ent2.3.1.xls
12. http://www.nsi.bg/ORPDOCS/ICT_Ent2.3.2.xls

EFFECTS OF THE GLOBAL INTERNET SPACE ON CULTURAL IDENTITY

Tanya Parusheva

University of National and World Economy – Sofia
t_parusheva@abv.bg

Abstract: Scientific report analyzes and discusses the effects of the global internet space on cultural identity. Attention is focused on the revolution of information and communication technologies as one of the main drivers of globalization in tourism. The utility of global distribution systems Sabre, Galileo, Amadeus and Worldspan is presented. The focus is placed on the new technologies that allow offering more flexible tourist product. The rapid dissemination of information reservation system in Europe is considered. The importance of electronic pictures and the ability to visualize images of tourist products is highlighted. An increase in the use of online reservation services via internet or intranet of companies specialized in tourism is found. The impact of globalization on cultural identity is thoroughly examined. The means by which to carry out cultural expansion - film, TV, music, internet are analyzed in detail. The opinion of anti-globalists that cultural globalization leads to a loss of national cultural values and creates a struggle for the strengthening of local cultures is presented. The author argues that the future development of each culture and the preservation of cultural identity are vital. The opinion that global internet space has a significant impact on the promotion of the cultural identity of the peoples of the world is summarized.

Key words: information and communication technologies, information reservation systems, Internet, globalization, cultural identity.

We live in a world of rapid change brought by the globalization. It is absolutely one of the most important economic, political and cultural processes that can not oppose any country. Cultural wealth of nations, have been accumulated over millennia. The global internet space can help the promotion of cultural identity.

Scientific report aims to analyze and discuss the effects of the global internet space on cultural identity. There is an opinion that cultural globalization leads to a loss of national cultural values and creates a struggle for the strengthening of local cultures.

The revolution of information and communication technologies

Information is the soul of the travel industry. It helps the relationship between tourists, tour operators, travel agents and providers in the tourism industry. Information is the basis that links the different producers in the tourism industry (cruise companies, airlines, hotels, tour operators, travel agents, transport companies and numerous other vendors). Information technologies are very important for the processing and dissemination of information effectively in a timely[1].

[1] Панджерова, А., Електронният туризъм – иновативен подход за малкия и среден бизнес в туризма, Дванадесети национален симпозиум „Качество, конкурентоспособност, устойчиво развитие", Университетско издателство „Стопанство", С., 2007, с. 117-122.

One of the main factors associated with globalization in tourism is the revolution in information and communication technologies. The largest and most important systems in the tourism sector are the **reservation systems** as "a central database, updated periodically, that is available to defined number of people". They appear as dominant technology[2].

Information reservation systems (CRS) have been developed by the airlines to control the availability of supply. Today there is a radical change in the distribution systems of global travel and increased competition among major airlines. They are trying to expand and promote the distribution of their products, developing global information reservation systems at the regional level.

Major **global distribution systems** (GDS) are **Sabre, Galileo, Amadeus** and **Worldspan**[3]. The utility of global distribution systems is presented in Figure 1.

<table>
<tr><td>Airlines →</td><td rowspan="7">Computer reservation systems →</td><td>→</td><td>Tour agents</td></tr>
<tr><td>Hotels →</td><td>→</td><td>Corporations</td></tr>
<tr><td>Rent-a-car →</td><td>→</td><td>Persons</td></tr>
<tr><td>Vacations/tours →</td><td rowspan="5">Internet</td><td rowspan="5">↑ →</td></tr>
<tr><td>Trains →</td></tr>
<tr><td>Cruises →</td></tr>
<tr><td>Ferries →</td></tr>
</table>

Figure 1. The utility of global distribution systems
Source: http://www.amadeus.ru/news/TravelTribes.htm

Sabre is the most popular system among travel agents in the United States. **Amadeus/System** one is the first in Europe. **Galileo** can boast of world performance. All four are used worldwide.

Changes that occur in the field of tourism and the increasing number of long distance trips are a trend that is manifested in terms of **tourism trips flat price**. New technologies in the tourism industry allow manufacturers to offer **more flexible products** at competitive prices[4].

Dissemination of information reservation system (CRS) in Europe

We live in a time when information technology and reservation systems (CRS), in particular, are going too fast. In Europe in recent years the number of travel agents, equipped

[2] http://www.papers.ssrn.com
[3] http://www.amadeus.ru/news/TravelTribes.htm
[4] Парушева, Т., Влияние на глобализацията върху поведението на потребителите, Издателство „Авангард Прима", С., 2010, с. 31-37.

with information reservation systems (CRS) has increased rapidly and their distribution increased from 80% to 90% in many countries[5].

Travel Agents use global distribution systems for information and reservations for air travel and accommodation. Some global distribution systems (GDS) have **electronic images** of their terminals. This allows tourists and travel agents to visualize images of tourist products.

Information reservation systems (CRS) are known in the business travel agency specializing in **leisure products**. Therefore, there is a high prevalence rates in all travel agencies in the larger European markets.

The role of the Internet

Internet is a network of the networks. Two of the most used functions are **electronic mail (e-mail)** and the **World Wide Web**. Nowadays, every family around the world daily gains personal computer. Everyone has access to multiple information systems.

The number of Internet users is increasing very rapidly. The implementation of the survey Nua results counting 90.9 million people connected to the Internet worldwide, including 54 million in the U.S. and Canada, 20.2 million in Europe and 8.6 million in Japan[6].

Tourism is popular on the Internet. The popular **search engine Yahoo** offers numerous **tourist sites** each of which contains a considerable amount of information. The use of **online reservation services via internet or intranet** various companies specializing in tourism is increased. Co-enterprise Worldspan/Abacus and software giant Microsoft have Internet travel reservation system called Expedia, which is truly remarkable.

Impact of globalization on cultural identity

The process of globalization has sharply exacerbated the problem of national cultural identity. Today, due to globalization, it is one of the most important issues that worry cultural anthropologists, politicians, religious leaders and members of progressive-minded science.

There are quite a few posts that criticize globalization. According to **anti-globalists**, rather than a means of enrichment and a source of replication of each culture with others, globalization becomes a form of **depersonalization** of all cultures[7].

Nowadays a process of transition from national to **global culture** whose primary language is the English. The U.S. dollar is used in many countries of the world. Mass Western culture rapidly penetrates into the lives of many countries.

Globalization lowers the **status of national language**. It establishes English as the only means of intercultural interaction, although it is mother language of about 400 million people. Now, a lot of books, newspapers and magazines are issued in English. Over 80% of materials on the **Internet** are English texts. The same can be said for audiovisual productions in the network, which is almost entirely created by English-speaking authors. Knowledge of

[5] http://www.europa.eu/areas

[6] http://www.businessline.com

[7] Парушева, Т., Глобализация в туризма, Издателство „Авангард Прима", С., 2010, с. 314-328.

English is always a requirement for those who want to find work in large companies, banks, insurance companies and others.

Experience shows that other languages, including those who are the official languages of the United Nations, gradually lose their significance. The range of their distribution gradually narrows actively used vocabulary is getting poorer and speech and written texts are saturated with **substitute language**.

Today **movies** are showing on the screen simultaneously in many countries worldwide. **Books** are translated and became popular among readers from different countries. Huge role in **cultural globalization** plays Internet penetration. Moreover, every year more and more popular is gained international tourism.

In the process of globalization disappears diversity of **cultural interaction**. Along with the process of deepening and widening of the dominant type of interaction between different cultural traditions is expanding. It provides the forcible introduction of the **value system** of a culture of values from **another culture**.

In recent years, massive saturation flows of space with different cultural patterns of American culture. So quite often, globalization is equated with **Americanization**. This identification is linked with efforts in the XXth century, the influence of the United States.

Today, Hollywood places on the screen most of the movies that attend international film poster. According to statistics, over 85% of the most popular movies are American (and in the UK, Brazil, Egypt and Argentina - 100%). E-mail and global network allow the U.S. to dominate the global flow of information and ideas.

Emerge from the U.S. **world brands** - Microsoft, Intel, AMD, Coca-Cola, Procter & Gamble, Pepsi and many others. McDonald's has become a **symbol of globalization**, because of its global distribution. Comparing prices of hamburgers "BigMac" from the local McDonald's restaurant in the countries of the world, the magazine "The Economist" analyze the purchasing power of different currencies - the so-called "Big Mac Index".

We can say that globalization is associated with Western culture. In summary, the means by which to carry out **cultural expansion** are:

- Cinema;
- TV;
- Music;
- Internet.

The process of globalization creates access to information on the Internet. Each member was introduced primarily to the views presented by the intellectual and political elite of the United States. All this leads to the fact that in a globalized world disappears **diversity of national cultures** and the dominance of Western culture.

The author of this report argues the proponents of globalization who opined that all modern processes and related negative phenomena caused by cultural globalization have a natural character. Globalization can cause homogeneity, but the alignment of the world will always enrich and preserve cultural identity.

Conclusion

All cultures can survive intact. Massive destruction of the leading crops would be disastrous. Future development of each culture and the preservation of cultural identity are

vital. Global internet space has a significant impact on the promotion of the cultural identity of the peoples of the world.

References

1. Панджерова, А., Електронният туризъм – иновативен подход за малкия и среден бизнес в туризма, Дванадесети национален симпозиум „Качество, конкурентоспособност, устойчиво развитие", Университетско издателство „Стопанство", С., 2007, с. 117-122.
2. Парушева, Т., Влияние на глобализацията върху поведението на туристите, Издателство „Авангард Прима", С., 2010, с. 31-37.
3. Парушева, Т., Глобализация в туризма, Издателство „Авангард Прима", С., 2010, с. 314-328.
4. http://www.amadeus.ru/news/TravelTribes.htm
5. http://www.businessline.com
6. http://www.europa.eu/areas
7. http://www.papers.ssrn.com

THE CHANGES IN LEISURE CONSUMPTION INDUCED BY THE USE OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Maya Tsoklinova

University of National and World Economy – Sofia
malenna85@yahoo.com

Abstract: The main purpose of scientific paper is to discuss the importance of traditional information and communication technologies for utilization of leisure time, and to analyze their impact on consumer behaviour. Special emphasis has been placed on customer preferences to actions for utilization of leisure time, which can increase the effectiveness of advertising and consumer loyalty. Traditional leisure activities such as reading, gardening and travel, have entertained the society for millennia. The new technologies have profoundly changed the way people consume their leisure time. It is interest also how traditional information and communication technologies for leisure utilization impact on consumer behaviour. Special attention has been placed on the role of effective advertising of tourist destinations in the company's website that promotes utilization of leisure time through tourism. The role of firms that they can improve their marketing by segmenting the leisure activities and connect these segments based on demographic and behavioural variables has been highlighted. Information and communication technologies generate new activities for utilization of leisure time. They create possibilities for reallocation of time and facilitate the practice of existing activities. It should also be noted that the increased use of information and communication technologies is growing in parallel with the increase in leisure time.

Key words: information and communication technologies, leisure time, changes in consumer preferences, new activities in leisure time.

Traditional recreational activities such as reading, gardening and travel entertain society for millennia. Moreover, all the technologies such as movies, radio, television, Internet and other information and communication technologies change the way people spend their leisure time.

Information and communication technologies find application both in private consumption of individual and business practices to stimulate consumer demand. Advertisers could attract customers interested in travelling if they have a strong interest in cinema. Destinations can be promoted by movies and films in tourist websites. Companies could improve their marketing through customer segmentation in recreational activities and linking these segments on specific demographic and behavioural variables. Customer-centric approach provides a competitive advantage to companies since products and processes become more standardized.

Impact of ICT on the consumption of tourist activities

ICT such as internet and mobile phones affect many aspects of daily life. These technologies affect leisure activities. Changes they bring are reflected in social interaction and choice of leisure activities. The practice of **traditional activities** such as games, writing letters or watching TV is changing. Information and communication technologies generate

appearance of **new leisure activities**, reallocate time need for realization of other leisure activities and facilitate the practice of existing activities. In addition, the importance of related ICT activities increases with the available time. With rising life standards, people tend to have more free time. As a result, people retire sooner or can work part time. Demand for leisure activities and the role of related ICT leisure increases with economic prosperity[8].

The additional effect of ICT on leisure can be considered about the **creation of new media**. Most new media do not arise spontaneously or accidentally. Their emergence and coexistence with old media generally lasts for decades[9]. So the old media are changing and sometimes disappear in response to new media. For example, radio survived after the advent of television and has become a mass technology. The process of adapting the consumption of leisure activities in the emergence of new technological developments is especially interesting.

Motivation to participate in leisure activities and ICT

Based on the motivation theory, some authors argue that consumers participate in leisure activities to achieve specific physical or psychological purposes. For example, some swim to keep fit or to increase their confidence. On this basis, some authors consider experience in utilization of leisure activities as a psychological outcome. People consume activities during their leisure time in order to **generate experiences** such as holidays for meeting with foreign cultures[10]. Studies of leisure time focus on the concept of motivation and behaviour and consider them as a function personal ability. Ryan learns leisure motivation scale of Beard and Ragheb, find four motives determinants of **satisfaction** gained from the leisure activities. The first motif is intellectual stimulation, for example, training or explore new cultures. Secondly, the scale is based on social causes such as building or strengthening relationships. Third, users engage in leisure activities for competing and defying themselves. Finally is the incentive to avoid overly stimulated life situations. These four motivational components influence the selection and use of new media[11].

Demographic and psychographic characteristics determine the use of new technologies for work or play. For example, people use the Internet at work and at home, to look for information about products and online shopping because of **convenience** and **time savings**. Also, they use the Internet for entertainment. Nearly one of three Internet users, mostly young men, regularly is going online for utilization of his leisure time. The Internet offers information, communication and recreation. Today, consumers spend moiety of their time with online media[12].

Moreover, the media facilitate social interaction. Within a household, families can discuss the specific content or play computer games. Outside the home, interactions with external partners is conducted primarily through the Internet, by mail or online communication. As media become entertainment oriented, there is a visible change in the consumption pattern of leisure activities. Some consumers prefer **advanced media technologies**, from improved

[8] Mokhtarian, P. L., Salomon, I., Handy, S. L., The impacts of ict on leisure activities and travel: A conceptual exploration, Transportation 2006, pp. 263-289.

[9] Ali, A., Frew, A., ICT and its role in sustainable tourism development, ICT in Tourism 2010, Springer, pp. 479-482.

[10] Парушева, Таня Петрова, Културният туризъм: алтернатива за глобализация, Издателство „Авангард Прима", София, 2009, с. 223.

[11] Ryan, Ch., Recreational Tourism: Demand and Impacts, Channel View Publications, 2003, p. 358.

[12] Пенчева, А., Управление на инвестициите в туризма, Авангард Прима, С., 2008, с. 53.

programming options to more traditional social media[13]. The following table shows us ratios in consumer preferences, in terms of utilization of leisure activities. There are distinct ICT related activities and traditional activities for utilization of leisure time, according to the preferences of the four different groups.

Table 1.

Preferences of leisure activities

	Traditional families	Internet consumers	Young people	Retirees
ICT-MEDIATES LEISURE ACTIVITIES				
Broadband access	48 %	63 %	58 %	61 %
Internet skills	65 %	84 %	80 %	24 %
Internet use	52 %	96 %	83 %	17 %
DVD watching	46 %	87 %	76 %	25 %
Watch DVD's	Evenings (weekend)	Everyday evenings	Anytime	Anytime
Cable TV	26 %	33 %	33 %	32 %
TRADITIONAL LEISURE ACTIVITIES				
Cinema	83 %	83 %	84 %	75 %
Travel	66 %	57 %	57 %	63 %
Literature	53 %	41 %	44 %	48 %
Sport	32 %	35 %	41 %	38 %
Health and fitness	45 %	26 %	39 %	35 %
Dining out	57 %	40 %	44 %	42 %
Museums	25 %	18 %	20 %	21 %
Food and wine	47 %	32 %	37 %	40 %
Gardening	23 %	20 %	19 %	30 %

Source: Zorn, S., Lee, R. Y., Muprhy, J., Marketing Implications of Traditional and ICT-Mediated Leisure Activities, Taylor and Francis, 2012, p. 329-341.

Traditional activities for utilization of leisure time, which include leisure travel, report the highest levels of consumption in families that have a higher disposable income. Highest level of preference of tourist activities for leisure time is measured in the latter group - pensioners. This group has a steady income and enough free time to plan the journey.

Changes in leisure induced by ICT

The advent of ICT's in all aspects of human life, and the impact of ICT's on personal choice and consumption if leisure activities are often discussed problem. Most studies are examined the impact of ICT on binding (work) and support activities (e.g. shopping).

[13] Tsoklinova, M., The role of information and communication technologies for development and organization of leisure activities, International conference "Application of information and communication technology in economy and education", University of National and World Economy, Sofia, 2011, pp. 699-706.

Discretionary activities for leisure activities have been widely studied topic. The study about the impact of ICTs on the consumption of tourism activities deserves its attention.

Trichotomy of Reichman considered personal activities that are built on individual motivation or objectives for inclusion in a specific activity, which can be earning money, satisfaction of psychological needs, entertainment or recreation[14].

- **Obligated activities** include work or work activities (for sustenance), which provide an economic basis for realization of personal activities.
- **Maintenance activities** include, purchase and consumption of goods and services that aim to meet individual family or physiological needs (shopping, banking, medical services) or biological needs (sleep, hunger, thirst, personal care) and obligations for membership to a community (housework, etc.).
- **Leisure activities** include discretionary activities (visits to restaurants, fitness, watching TV, etc.).

There is enough statistical information which presents data on cultural differences that are observed in the utilization of leisure time. People allocate about one third of the available time for leisure time. A growth in the amount of activities that take place during leisure time is observed.

Leisure trips cover a significant part of everyday life. As can be seen in the following figure, 27% of daily trips are associated with social and recreational activities, such as visiting relatives and friends, watching movies, going on holiday or take part in sports activities.

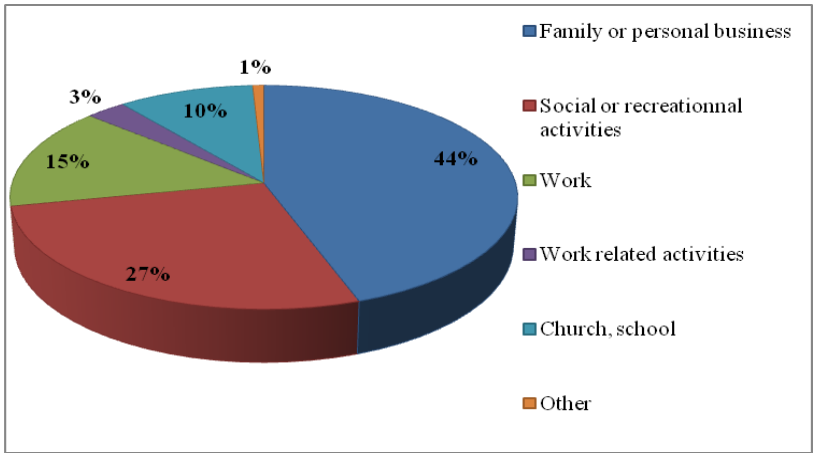


Figure 1. Purposes of personal travel
Source: www.bts.gov/publication

The study of the impact of ICT on leisure and tourism is relevant for the implementing of **policy problems** such as congestion, pollution, urban planning.

What is leisure for one is not leisure for other. For certain people, under specific circumstances, all activities for utilization of leisure time can be regarded as mandatory or supporting. Most studies of the behaviour during the trip are focused on the mobility of certain activities in relation to spatial flexibility.

Table 2.

[14] Reichman, S., Travel Adjustments and life styles: a behavior approach, Massachusetts, 2003, p. 172.

Relevant dimension of leisure

Group	Dimensions
Location	Location (in)dependence, mobility-based vs. stationary.
Time	Time (in)dependence, planning horizon, temporal structure and fragmentation, possible multitasking.
The social context	Solitary vs. social activity, active vs. passive participation.
Traits intrinsic to the activity	Physical vs. mental, equipment/medical (in)dependence, informal vs. formal arrangements required.
Benefits and costs	Motivation, costs.

Source: Mokhtarian, P. L., Salomon, I., Handy, S. L., The impacts of ict on leisure activities and travel: A conceptual exploration, Transportation 2006, pp. 263-289.

Doherty offers remarkable features of activities that explain complex tourist behaviour. Spatial and temporal flexibility, considered as dimensions that can explain the complex behaviour of the individual are discussed in various studies.

The concept of interaction between ICT and leisure is considered by many authors. ICT in most cases changing tourist behaviour, but there is still uncertainty about these changes. So Kupam and Pendaya offer a conceptual framework for studying the interaction between socio-demographic characteristics, participation in activities and tourist behaviour[15].

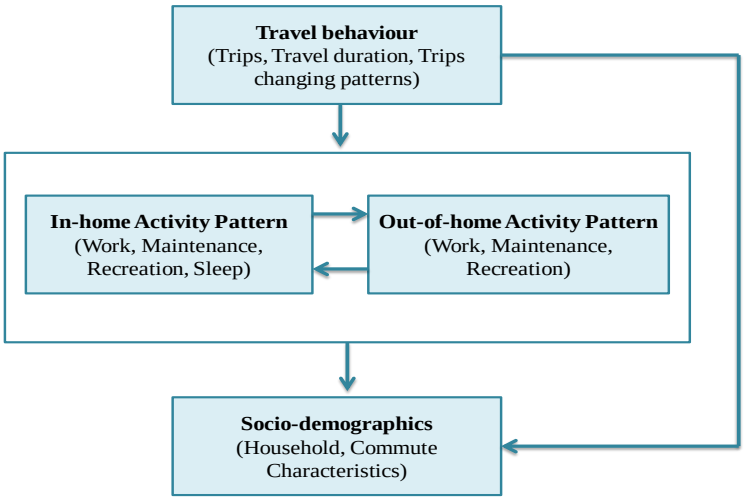


Figure 2. Framework of leisure activities

Source: Kitamura, R., The Expanding Sphere of Travel Behaviour Research, Emerald Group Publishing, p. 317.

[15] Kitamura, R., The Expanding Sphere of Travel Behaviour Research, Emerald Group Publishing, p. 358.

Handy and Yantis argue that there is no connection between going to the movies and travelling. These two activities are not interchangeable, as ICT is still not as advanced as to provide real satisfaction such as travel.

To clearly define the changes in consumption of leisure activities influenced by ICT, it is necessary to classify them appropriately. Free time can be described FTAM model (Flexibility-Time-Allocation). Spatial flexibility of leisure activities is recognized the fullest when performing activities may occur anywhere. Time flexibility is determined by the ability to perform activities anytime. Overall flexibility of the model is determined by the ability to carry out activities in different ways. The flexibility of leisure is represented graphically in the following figure.

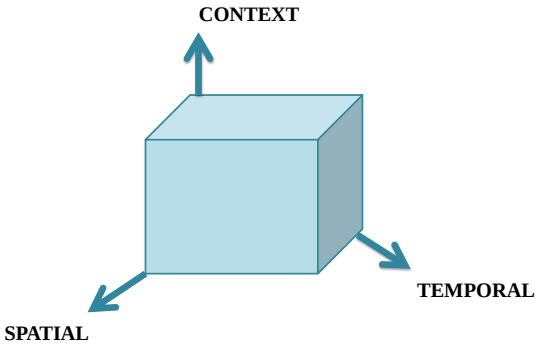


Figure 3. Three-dimensional flexibility of an activity
Source: *Andreev, P., Pliskin, N., ICT impacts on leisure activities and leisure-related travel, University of Negev, Israel, 2008, p. 45.*

In general we can distinguish activities for leisure activities, as a distinctive feature in their performance. They can be:

- Technology based.
- Depending on the location.
- Depending on the mode of transportation.

Changes in consumption patterns can be studied on the basis of certain factors. Their relationship is presented in the next figure.

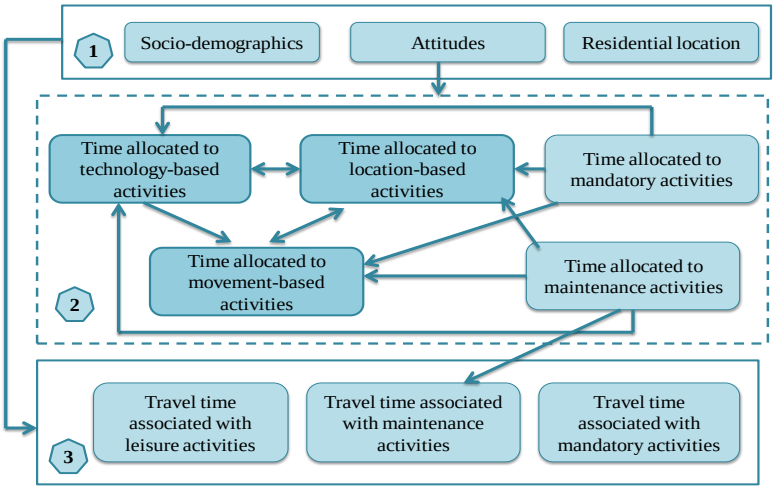


Figure 4. Flexible Time Allocation Model

Source: Andreev, P., Pliskin, N., ICT impacts on leisure activities and leisure-related travel, University of Negev, Israel, 2008, p. 45.

The proposed model considered choice of activities based on socio-demographic categories. Each of the activities, however, there are certain characteristics of consumption, which significantly affect the development of ICT. As a result, permanent changes occur in the selection of tourist activities.

Implications

- Companies could improve their marketing through customer segmentation in recreational activities since customer-oriented companies gain a powerful competitive advantage.
- ICT in most cases change tourist behaviour and lead to increased consumption of activities for leisure activities.
- The impact of ICT on the consumption of tourism activities is expressed in an increase in the amount of activities that take place during leisure time.

Conclusion

In conclusion, we can determine that the ICT impact on the way people utilize their leisure time and encourage the consumption of more tourist activities. ICT change the motivation and the means by which people manage their time and have access to activities for leisure activities.

References

1. Парушева, Таня Петрова, Културният туризъм: алтернатива за глобализация, Издателство „Авангард Прима", София, 2009, с. 223.
2. Пенчева, А., Управление на инвестициите в туризма, Авангард Прима, С., 2008, с. 53.
3. Ali, A., Frew, A., ICT and its role in sustainable tourism development, ICT in Tourism 2010, Springer, pp. 479-482.
4. Andreev, P., Pliskin, N., ICT impacts on leisure activities and leisure-related travel, University of Negev, Israel, 2008, p. 45.
5. Kitamura, R., The Expanding Sphere of Travel Behaviour Research, Emerald Group Publishing, p. 317.
6. Mokhtarian, P. L., Salomon, I., Handy, S. L., The impacts of ict on leisure activities and travel: A conceptual exploration, Transportation 2006, pp. 263-289.
7. Reichman, S., Travel Adjustments and life styles: a behavior approach, Massachusetts, 2003, p. 172.
8. Ryan, Ch., Recreational Tourism: Demand and Impacts, Channel View Publications, 2003, p. 358.
9. Tsoklinova, M., The role of information and communication technologies for development and organization of leisure activities, International conference "Application of information and communication technology in economy and education", University of National and World Economy, Sofia, 2011, pp. 699-706.
10. Zorn, S., Lee, R. Y., Muprhy, J., Marketing Implications of Traditional and ICT-Mediated Leisure Activities, Taylor and Francis, 2012, p. 329-341.
11. www.bts.gov/publication

Intelligent growth - an untapped opportunity for Bulgarian business organizations

Tsvetana Stoyanova

Management Department of UNWE – Sofia, Bulgaria
e-mail: tzvetana_stoyanova@abv.bg

Abstract. Research and innovation can help Bulgaria build the value chain in sectors where the country has a comparative advantage, and increasing the share of high tech products for export, thereby strengthening the competitiveness of business. It is necessary research to be more closely associated with the companies through the development and commercialization of R & D, with an active patenting and licensing, formation of companies and contract research. It is urgent to stop the erosion of technological and scientific expertise through strategic support of research and innovation. More and more consistent investment in public and private R & D could resume the innovative potential of the country and it could increase its competitiveness. All this defines the purpose of this report, namely to explore the three main factors forming the "intelligent growth" - education, innovation and the digital society to enhance business competitiveness and the economy as a whole.

Key words*:* intelligent growth, innovation, business competitiveness.

**

The challenge for Bulgaria in the field of competitiveness is to achieve intelligent growth meaning that scientific research and innovations shall become some of the main engines of economic growth. Bulgaria's competitiveness has increased in recent years, but not enough to overtake the other new member states of the EU or to keep pace similar to that of the emerging economies in Asia and Latin America. Throughout the last decade Bulgaria's export increased in absolute terms and per capita, and in 2008 it accounted for nearly 60% of GDP. The export basket of Bulgaria is still dominated by traditional products and mature industries. Today, the share of high-tech exports to Bulgaria is only 3% of total exports, which is significantly below the average for the EU-27 (16%). Innovation could support sectors in which Bulgaria has a comparative advantage to upgrade their activities along the value chain and expand the base of high-tech products for export.

By the end of 2009 the innovation activity of the business in Bulgaria were realized with pace and structural changes which definitely did not satisfy and did not benefit for its competitiveness on the European market.For the same period, only 16 % of the enterprises in Bulgaria developed innovation activity, and by that index our country took last place among the member-states of the European Union. To compare with, in Europe 40 per cent of the enterprises invested in innovations.

As a share in the Gross domestic product (GDP) the investments in innovations in Bulgaria tend to decrease. In 2000 the percentage was 0,52 per cent, while at present it is - 0,48 per cent of the GDP. To compare with, the percentage in Sweden was 3,82 per cent, and in Germany – 2,51 %. Bulgaria had the lowest gross domestic product per capita in the European Union and the smallest share of investments in innovations. The share of scientists and engineers on the labour market in Bulgaria was 3% of the human resource in our

country. That share in Belgium was 7,9 per cent. Besides, it is typical for the Bulgarian companies that they compete with offering low-price production, not with innovations.

On the grounds of these general features, we examined one of the main factors of intelligent growth – innovation policy in order to research the potential and to reveal the innovative activity of the business of the South-West region for planning and to show to what extent organizations, creating innovations take advantage from the national and European means of support. For the purpose object of research were 150 organizations from the production sector, services, telecommunications and bank institutions. The analyses and assessments were based on 118 specific answers, or on 79% of the initially appointed organizations, which we believe is quite representative.

The interest and the information requested by us concerns: Innovation resources Funding, Expenditures on innovations, Introducing new products, Participation in joint companies, Factors for creating an innovation system, Products/services added value, Labour productivity, System for sharing knowledge in the organization.

For the elaboration of the methodical part of the research work we used the Likert scale, i.e. 5-rate grade. The defining of the innovation index of the Bulgarian enterprises we calculate like:

- Generalizing measure of the innovation activity on a company level, aggregating seven different types of innovations out of four types, realized by the enterprises (product, process, organizational and marketing), and their degree of novelty (new for the enterprise, for the market, or for the world). It takes values from 0 to 100, index 0 meaning that the enterprise has not innovated at all, and 100 that it has realized all types of innovations with maximal degree of novelty.

- Components of the innovation index of the enterprises are the product innovations, the process innovations and organizational innovations:

When researching question Group one, concerning the innovation activity of the questioned enterprises, we found out that innovations are directly dependent on a number of inside and outside factors: features of the market, on which they are working, possession of internationally accepted standards; applying the principles of strategic positioning; implied IT systems. The realization of the innovation process presumes serious advance preparation – choice of innovation project, supplying the necessary funding, training or hiring staff, reorganization of the company activity. As a compensation, the achieved results (new or improved products and processes) can turn into a base for the achievement of the company long-term objectives.(see Fig.1).

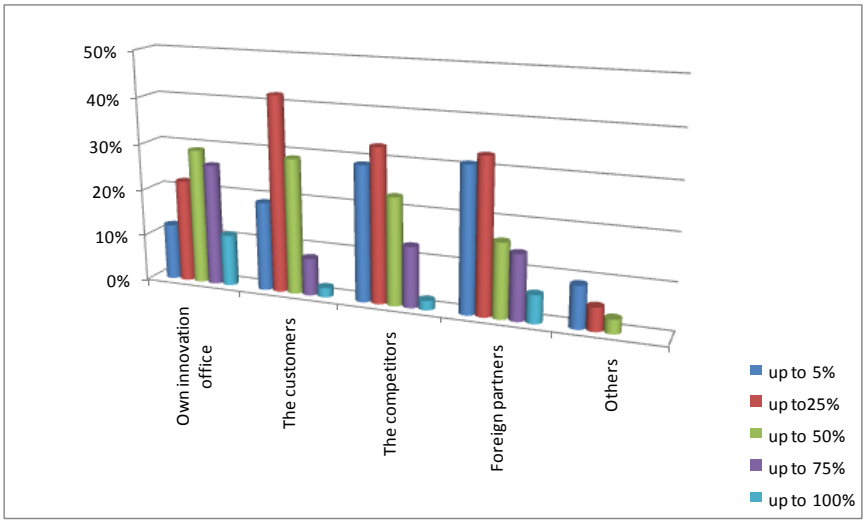


Fig. 1. Type and structure of the innovation resources in the researched companies

It is clear that almost equally the possible resources are assorted – own innovation office, customers, foreign partners, competitors. Only 12% of the questioned rely on their own innovation office for the complete elaboration of innovations.

Of crucial importance for the realization of innovation activity is the presence of enough in number and potential sources of funding. We define this resource of crucial importance for the success of the innovations. The sources of funding used by the companies in the SWRP reflect as a whole:

- ***general*** *innovation culture, expressed in:*
  - Lack of close connection between the parts of the regional innovation system and
  - Relying mostly on own forces on behalf of the companies when investing in product and process renovation;
- *environmental conditions for stimulating the creation and introducing new technical knowledge, which are function of:*
  - Unsuitable policy for funding and crediting, which is not consistent with the condition and possibilities of the business.

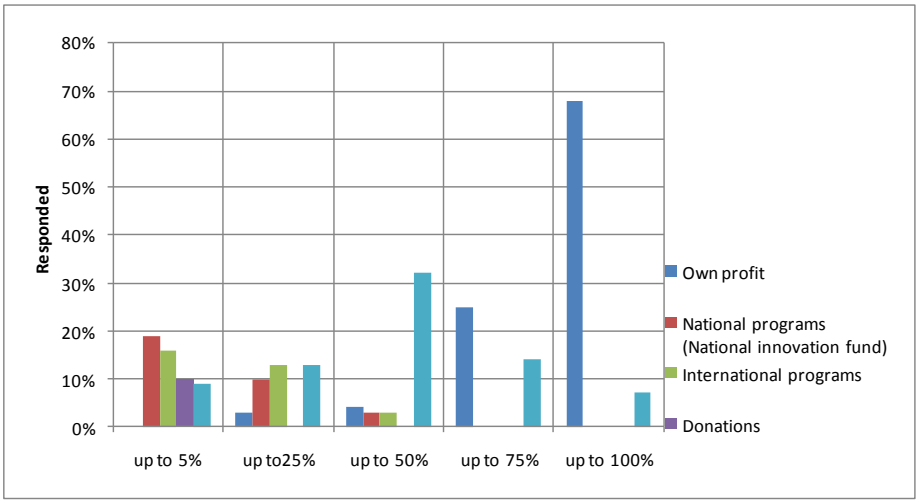


Fig. 2. Type and structure of the sources of funding of the innovation activity in the enterprises

As we can see from the data in Fig. 2, the enterprises mostly rely on their own funds and bank credits, and almost completely neglect the opportunities to participate in national and international programs. So 68% of the questioned rely on their own funding for the development of innovations, and for 32% half of the necessary funding comes from bank credits (outside sources). Only 12% of the questioned have taken advantage from national and international programs, and to the extent of up to 25% of the necessary resources.

The technological structure of the used resources for innovation activity definitely shows extensive approach and materializing the funds into long-term material assets. (Fig. 3)

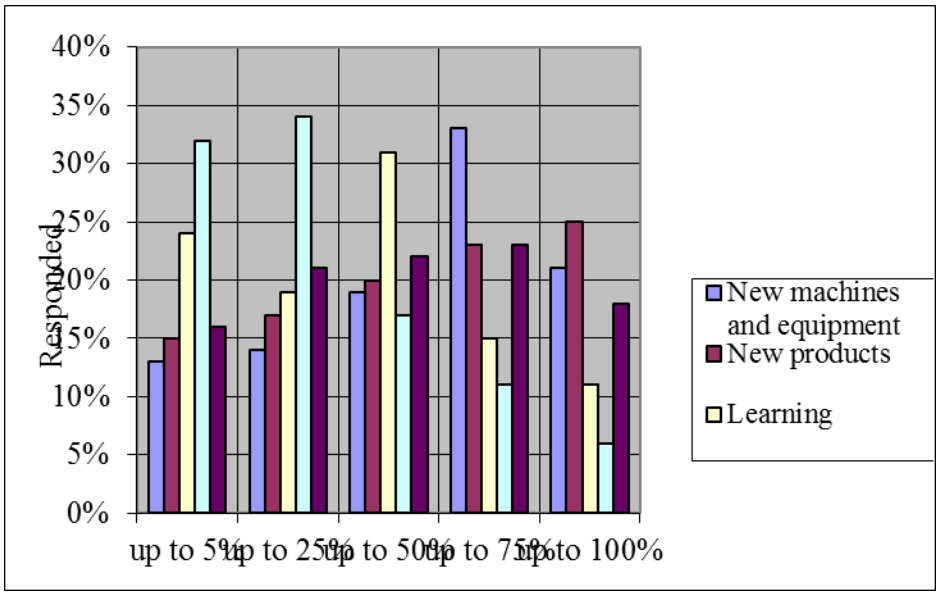


Fig. 3. Structure of the used financial resources for innovations

As we can see, over 50% of the enterprises management bind their expenses on innovations with buying new machines and equipment. They are followed by those who consider that expenditures on innovations are connected with new products (45%). This for sure makes the innovation activity ineffective, as it leads to recapitalization of the enterprise and, combined with not very high productivity, it reduces the efficiency and its competitive status.

The capability of the companies to create new products, and the skill to commercialize them successfully and on the international market too, is a base to achieve technological competitiveness. We realize that there is a restricted circle of economic units which have the necessary technological competences to make leading, in international aspect, product decisions, and experience in the area of introducing innovations. Here we have in mind equipment with modern technical specifications, protected technologies for the organization and realization of production processes, systems for the production of final products with constantly high quality.

We found out that the so called “Process innovations” are rare despite the obvious necessity for their realization. They are the ones that improve the environment for innovations in the company and ease the inside-company cooperation. This unfavorable situation can be seen from the data in Fig. 4.

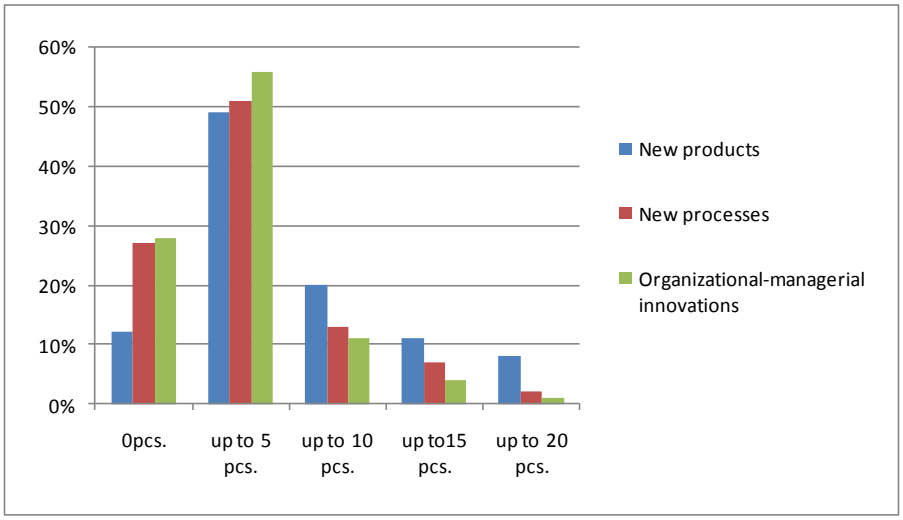
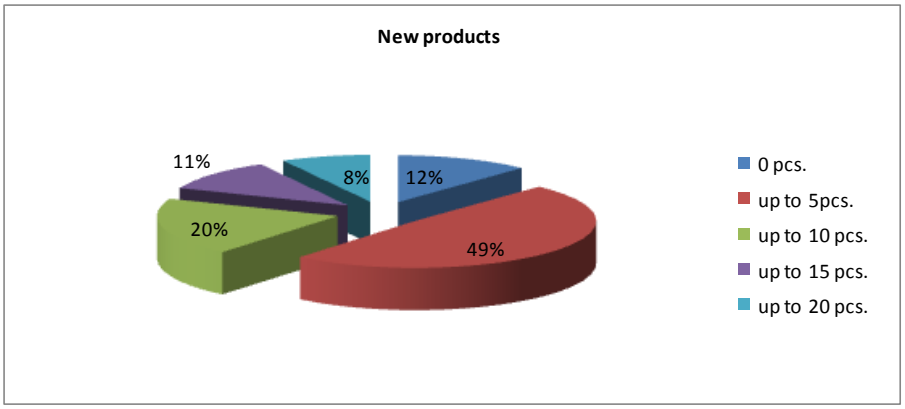


Fig.4. Type and technological structure of innovations

As we can see, in the last three years the researched enterprises have introduced at most 5 new products, processes and organization-and-managerial innovations, which is extremely low level of product innovation with increasing innovation demand on the different markets.

We can objectify the company policies by different types of innovations and assessment of their activity examining the following group of data (Fig.5).



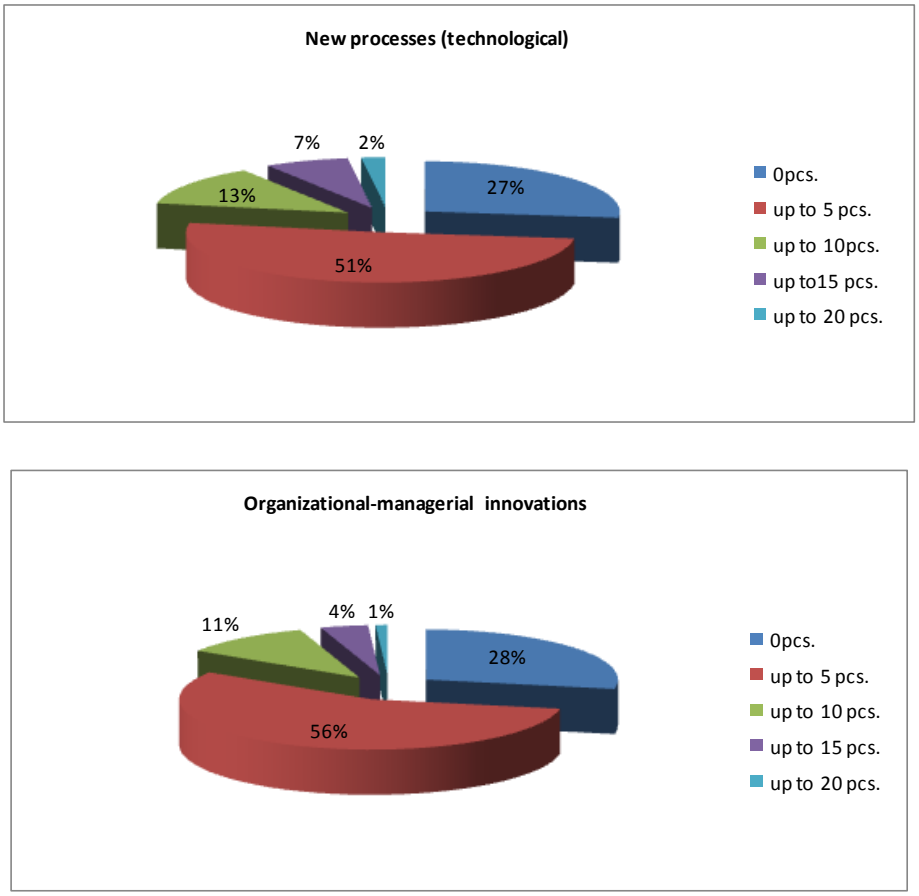


Fig.5. Dynamics in the structure of the basic innovations in the activities of the researched enterprises

Or, we can generalize in this part that:

- From 29 to34% of the enterprises make innovations every year, 90% of them have fixed, non-reducing budget for innovations, and half have increased it in a 2009 г. Compared to 2008. About 7 to 10% of the enterprises innovate sporadically.
- The average index of innovativeness of the enterprises, working mostly on the international markets, is three times higher compared to the ones working only on the local markets (situated within 30 км from the enterprise) and twice as high compared to the regional markets (situated within 100 км from the enterprise). Innovations are long-term engagement and require specific attitude, which cannot be created in a year or a few months, which is most often the range of the operative planning. According to the data the average innovativeness of the Bulgarian enterprises, the horizon of which is 3 years, is by 50 % higher in comparison with the enterprises with horizon up to 1 year.
- Enterprises which make long-term plans usually, on the base of the company experience and previous practices impose micro tends at the amount of the

innovation investments. This leads to the introduction of radical product and process innovations and to considerable changes in the users demand.

One of the questions of interest is how much the innovation activity should be of the type "Independent activity", or should be realized looking for support and partners by joining different business networks (clusters). As it is known, one of the possible alternatives here is to create mixed companies or different types of Joint Ventures. The data show that 78% of the questioned companies are not part of joint companies which would guarantee greater opportunities for know-how transfer, innovations and possibilities for knowledge management (see Fig.6).

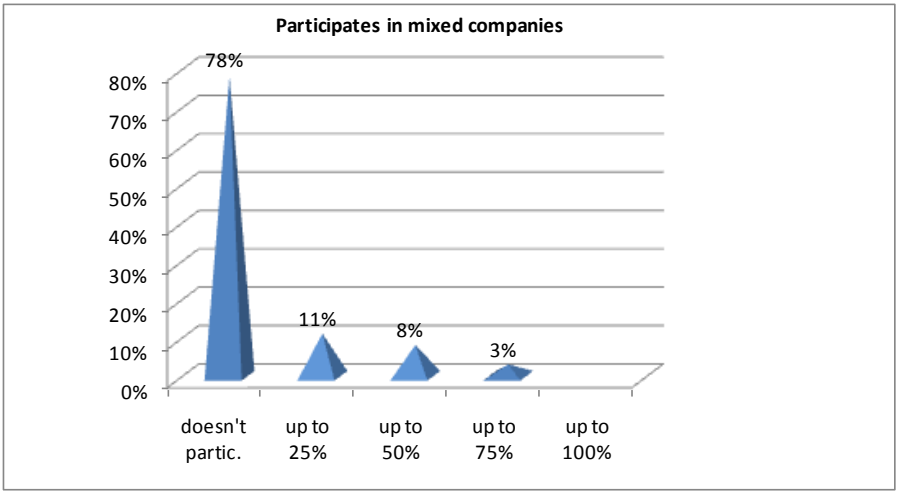


Fig. 6. Integrity in the innovation activity of the enterprises

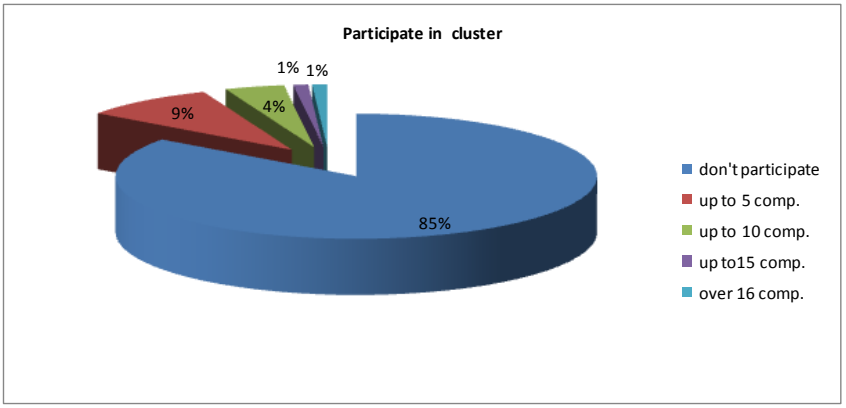


Fig. 7. Participation in innovation activity through "Clusters"

Looking for support and partnership in the innovation activity to a great extent is defined by the innovation potential of the enterprises and their competitive qualities. Here we have in mind generally the character and productivity of the inside-research activity and certification of activities, processes and systems of management in the organization. The

condition of these important sides of the competitiveness of the Bulgarian enterprises can be seen in Fig. 8 and 9.

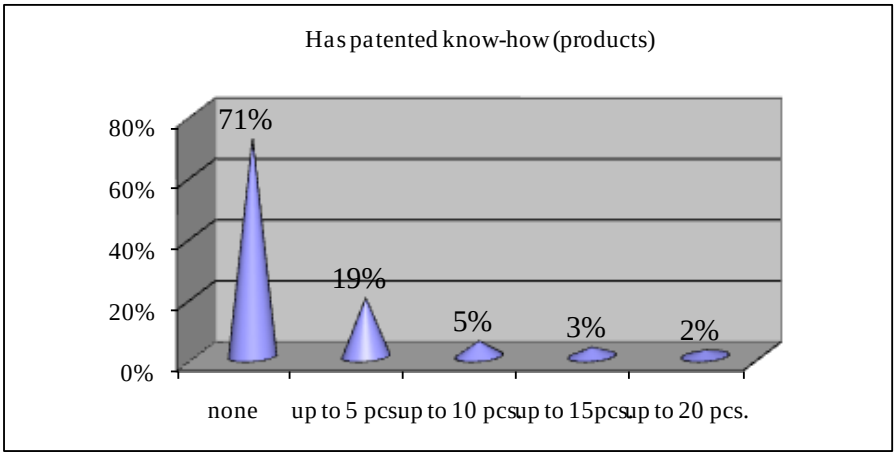


Fig. 8. Patented products in the organizations

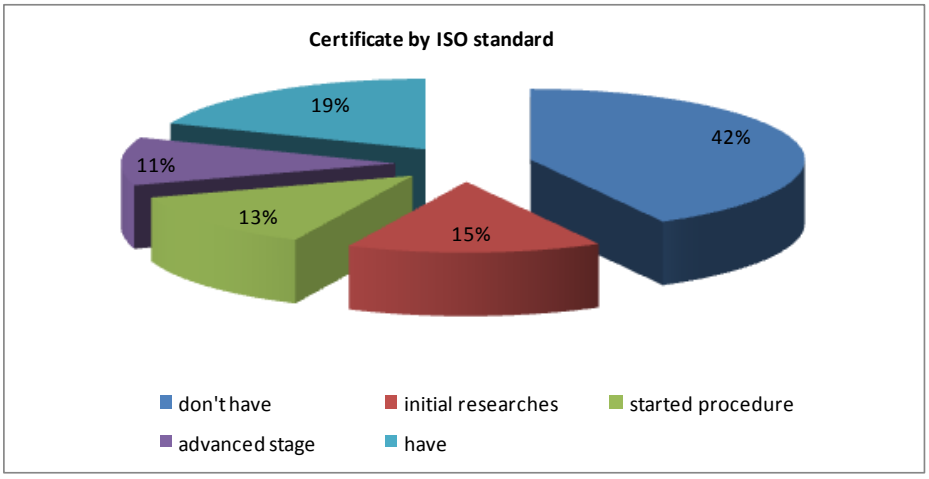


Fig. 9. Level of certified activities and processes

As we mentioned before, to obtain real increase of the competitiveness of the Bulgarian companies, it is necessary to really bind and balance the innovation targets with the strategies of the enterprises. Our research shows that the lack of strategy based on knowledge and innovations, is in the core of the lack of increase in the labour productivity and the amount of the added value. This shortcoming has not been overcome by the company management yet. (see Fig. 10)

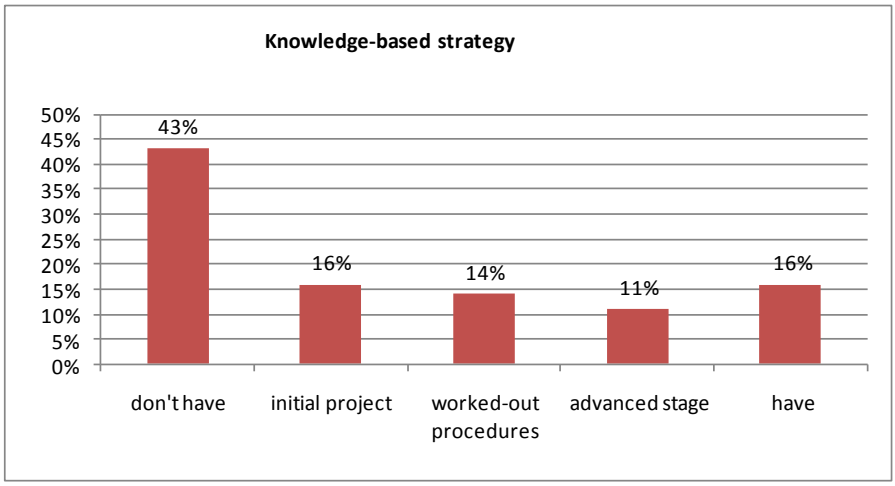


Fig. 10.Condition and presence of "innovation strategy"

On this base the opportunity to increase the competitiveness on the base of the innovation factors should strengthen the power of the approved scientific, together with the international organizations for this development (Fig. 11).

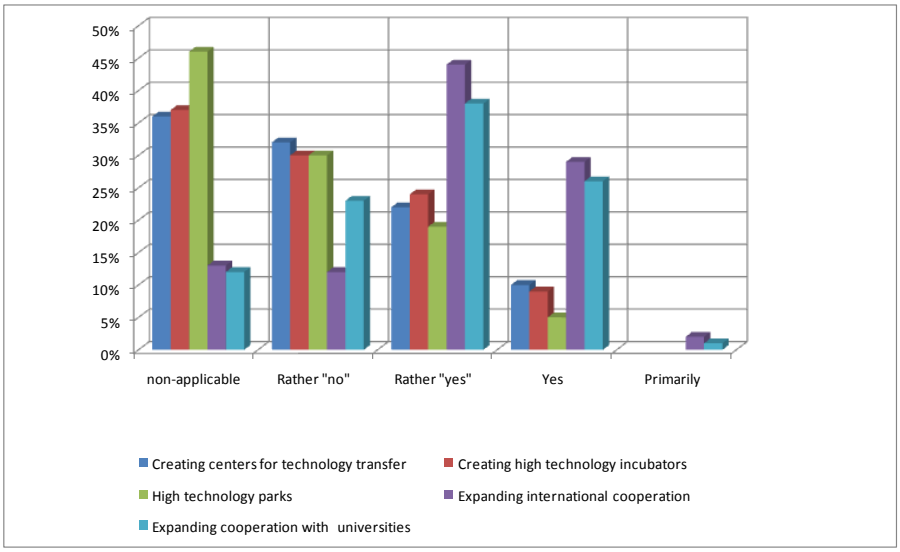


Fig. №11.Sources of innovation policy in the enterprises

As we can see, in our enterprises predominate the understanding that the expanding of the international cooperation and that with a university are basic factors for the improvement of their innovation system (75%). In spite of that, the effect from the innovations, expressed in the added value for about 25% of the enterprises, is highly appreciated, and concerning the labour productivity it is typical for 45% of the Bulgarian enterprises. (See the data in Fig.12 and 13)

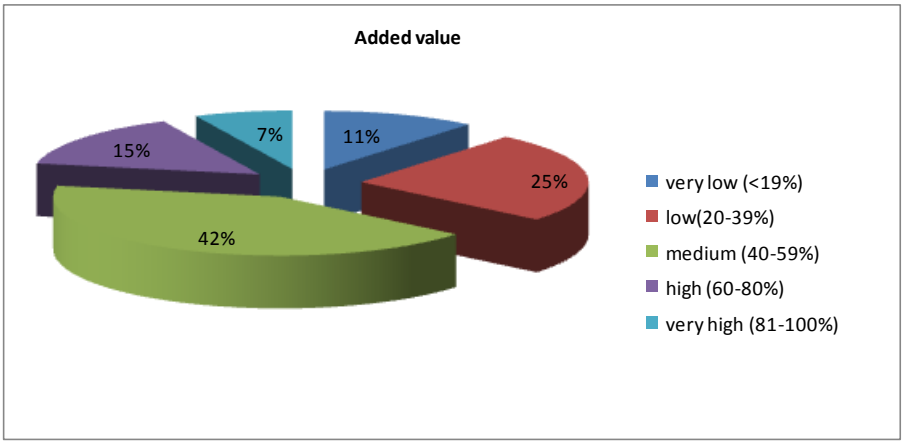


Fig. 12.Structure of the added value from innovations

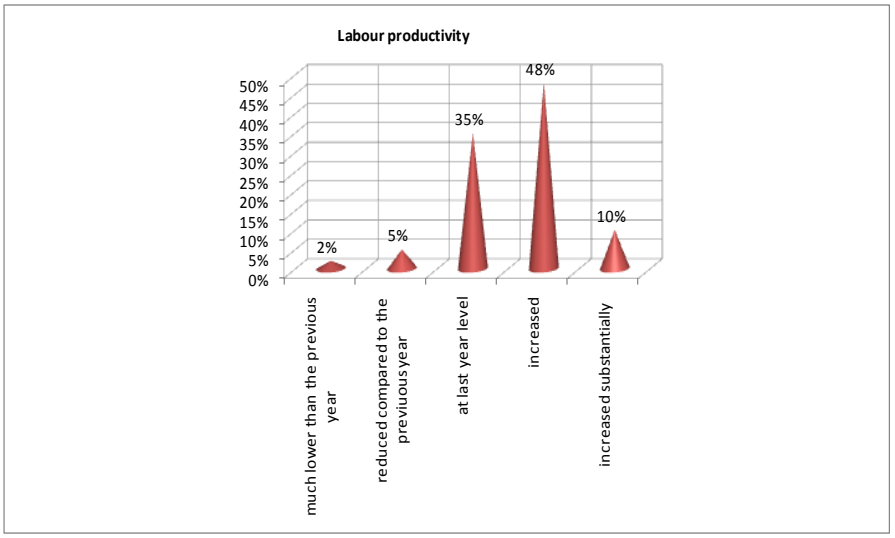


Fig. 13. Correspondence of innovations with labour productivity in the researched enterprises

***

As a summary of the results from the survey on the status and level of the components of intelligent growth in the Bulgarian economy and business, we may conclude that:

1. Big part of the questioned companies do not have strategy based on knowledge; The innovation activity is comparatively low. The respondents accept the system for sharing knowledge mostly using e-mail. The labour productivity is low in comparison with the average for the EU. Innovations are most of all connected with buying new machines, equipment and products, not with their creation. Joining clusters is not enough, which is an opportunity to incorporate the potential of the organizations, and achieve bigger synergetic effect.

2. To build up efficient strategy for knowledge management in the Bulgarian enterprises it is necessary to clearly identify its basic elements, sub-systems and factors, which are specific for each organization. "Mapping" the knowledge and the transfer of the best practices should be based on results from comparison with the branch leaders. The essence of forwarding the best practice is not to copy it, but adapt it to the specific requirements of the economic subject.

As a whole, on the base of conclusions about the innovation potential of the Bulgarian enterprises, we make the following recommendations:

In a mid-term period it is necessary to assure the following several policies:

- Stimulating enterprises to transfer technological, product and process innovations from leading scientific and research centers and business organizations in the country and abroad.
- Shift from the policy of budget funding of the educational and scientific and research organizations to a mixed type of market investment funding. The aim is for the business organizations to join efficiently business networks and clusters with high technological level.
- Improvement of the legislation, especially minding the outrunning integration of Bulgaria on the international knowledge economy markets.

The acting "standard" liberal economic policy should be completed with a complex of measures, directed to supporting the knowledge economy sectors: creation and distribution of innovations, ICT markets and the contemporary system of education. This requires:

- Vast increase in the absorption of EU funds and increase in the public spending on R & D, according to the capability of the state budget. Carrying out business in the well-known manner so far shall not lead to the required growth of investments in R&D.
- Adoption of new strategies that shall make innovations a driving factor for the export, for the establishment of jobs and for the economic growth.
- reform in the institutions and the forms of funding for science and education, providing reduction of the relative and absolute losses in them, as well as orientation toward future needs/ government demand/actual needs/ market demands;
- measures for the use of ICT in economy, the social area, for management and development of the information and telecommunication markets (creating infrastructure, i.e. personnel, legal, telecommunication infrastructure of the use of ICT by companies and citizens).
- Activating the participation in joint programming with the EU member states in order to achieve synergy.

Technological Challenges Faced when Organizing Accounting Operations

Borislav Boyanov

Abstract. This publication highlights the organization of accounting operations comprising a logically structured system of specific conditions and elements applied in the process of accounting reporting, aimed at obtaining high quality information about the enterprise's economic activity and achieved results. Depending on the researcher's point of view, different aspects of accounting organization are shaped within the accounting theory. To the author's opinion, it is most suitable to establish a normative and management aspect. The subject matter of this publication is oriented to one of the major groups of factors affecting the management aspect within the accounting organization – information technologies. Modern IT solutions approved the applicability of the so called "automated form of accounting". The body of this publication is presented through the prism of one of the most widely applied accounting software in Bulgaria – Microinvest Delta Pro. The publication considers the reporting process technology, its stages, used accounting registers, report generation and the principles, benefits and innovative functions of the accounting software.

Keywords: capacities, functions, software, accounting, organization

1 Introduction

The historical development of the economic science is an integral part of the societal progress. At a certain stage in the course of this process, the economic science has been divided into different disciplines. "The creation of accounting – the first form of identifying the economic sciences, and reporting knowledge of reality, respectively, puts the beginning of such division"[16].

Etymology of the term "accounting" originates from "account keeping" or an occupation involving trade and other accounts keeping in designated books within an enterprise, revising of such books, as well as preparation of statements for the financial position.

At present accounting should be considered as:

1. A system.
2. A process.
3. A knowledge area (science).
4. A profession (practice).

This publication highlights two of the accounting aspects – as an information system and purposeful human activity (practice).

[16] Trifonov, T., Accounting Theory: Philosophic and Economic Principles and Modelling, University Publishing House Stopanstvo, Sofia, 2008, page 13

With view of the systematic approach, accounting is considered **an information system**, which statically and dynamically monitors, measures, registers and analyses the enterprise's property and its formation sources.

Accounting **as purposeful human activity (practice)** means continuously developing system, a set of principles and formulations affected by the dynamically developing socio-economic processes.

2 Paper Preparation

2.1 Organization of Accounting. Forms of Accounting

The unity and sustainability of accounting as a system and as practice are due to the existing organization level. In this sense, the organization of accounting is a logically structured system of specific conditions and elements applied in the process of accounting reporting, aimed at obtaining high quality information about the enterprise's economic activity and achieved results. The organization elements of accounting may be systemized and summarized in the following four groups as shown on figure 1.

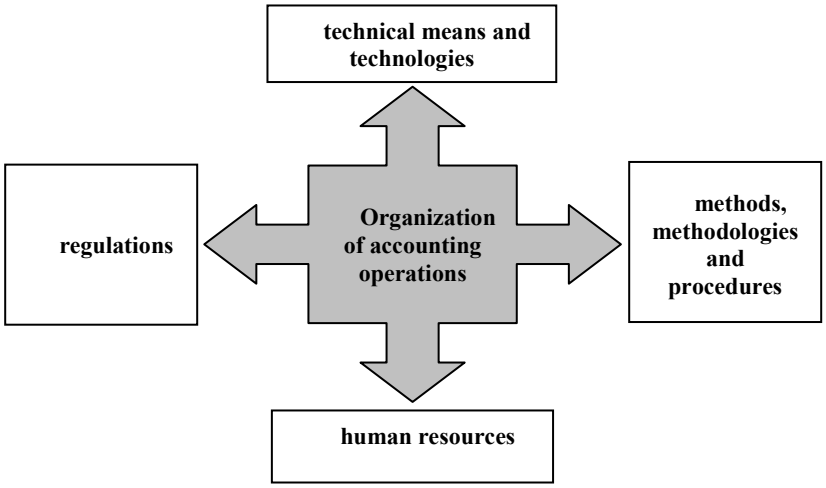


Fig. 1. Major elements of the organization of accounting operations

Organization of accounting should be considered in two directions:

First, management direction – it is relevant to the enterprise's organizational structure and the place of the accounting system; to the human resource management, to the management of accounting staff within the enterprise's accounting unit; as well as to the overall technology of the reporting process, to the applied methods, methodologies and procedures; to the use of the respective equipment, information and communication technologies (technical means and information carriers) for accounting information processing and analysis.

Second, normative direction – it is relevant to the adoption and application of specific restrictive conditions, norms, regulators. To this end, the organization of accounting has statutory limited boarders as provided for in the applicable legislation, accounting standards and internal rules for performing accounting operations within an enterprise.

The choice of an appropriate form of accounting is a key element within the management aspect of the enterprise's organization of accounting operations. **Such form of accounting comprises a set of approved and used accounting registers and the technique for fixing reporting data on such registers**.

The first historical evidence for the use of accounting forms dates back to the end of the 13th and the beginning of 14th century, when the scientific foundations of accounting are laid. During this period, the accounting policy uses the so called "previous single-entry Italian form". Afterwards, the accounting evolution and the development of accounting information fixing, processing and summarizing technique within specific registers result in the establishment and practical application of new forms of accounting. In chronological order, the most famous and widely applied forms of accounting since the 13th century until nowadays are as follows:

1. Previous single-entry Italian form of accounting.
2. New single-entry Italian form of accounting.
3. US form of accounting.
4. French form of accounting.
5. German form of accounting (South German and North German[17]).
6. Logismography.
7. Russian form of accounting.
8. Chess (matrix) form of accounting.
9. Memo-order form of accounting.
10. Entry-order form of accounting.
11. Automated form of accounting.

Regardless of the great variety of accounting forms and the reporting process automation level, each form should always consists of: a system of accounts and entries therein, a system of chronological and systematic (synthetic and analytical) entry and summary of reported data, and a system of current and regular identification of indicators that characterize the enterprise's property and financial position, the changes in cash flows and equity.

[17] See Belmer, F., Common Theory of Accounting, Third edition, Varna, 1932, page 587-590

The introduction of information and communication technologies in the process of accounting reporting allows to make accounting operations automatic. Thus, a new form of accounting is established and identified, which is called **"automated**". At present it is realized by means of computer information systems as a set of technical and software components.

The main principles of automated forms of accounting, which are also considered advantages, mainly refer to:

- Single entry of data reported from primary accounting documents
- Automatic creation of analytical and synthetic (systematic) accounting reporting registers;
- Automatic generation of chronological register of accounting entries;
- Automatic retrieval of information for any deviations from established norms and standards;
- Automatic generation of all required accounting registers, sample forms of financial, tax and statistic statements.

Contemporary technological challenges faced when organizing the accounting operations are expressed in decision making for introduction of specific accounting software. The selection of the most appropriate software runs through several phases related to the enterprise's scope of business, the economic grounds, the "benefits-costs" analysis, compliance with accounting and tax legislation, adaptation to the organizational structure of the accounting unit, defining the access rules, information security means, support of accounting software by the developers, quick and timely satisfaction of information needs of the enterprise's management, etc.

2.2 Technological Challenges Faced by Accounting Operations. Microinvest Delta Pro Accounting Software

The market of accounting software usually operates within the frames of national economies. The major factors restricting the globalisation of this market are due to the existing economic, political, cultural and other factors expressed at national level. There is a restricted circle of "big stakeholders" who have the resources and abilities available to adapt the products they offer in different countries and even in international aspect. Therefore, their clients comprise of large multinational corporations that have subsidiaries and divisions in a number of counties located at different continents. These corporations introduce not different accounting products but entire business-information systems used for automation of all resources management – tangible, intangible, personnel and financial.

Small and medium size enterprises in Bulgaria find it most economically and organizationally justified to implement accounting software. Varied double-entry accounting products are offered on the Bulgarian accounting software market. One of the most sellable products here is **Microinvest Delta Pro**, developed by Microinvest LTD. This is a new generation accounting system based on SQL platform which combines three major factors for success in every undertaking – speed, comfort and reliability. Thus, it ensures stable operation within a network and access via the Internet, which allows information real time entry to a single database from a few remote work places. The product is compatible to Windows XP\Vista\7 32-bit and 64-bit.

Microinvest Delta Pro is to the highest extent compliant with the requirements of the automated form of accounting. It complies with the Bulgarian accounting and tax legislation and features as follows

It allows organizing accounting operations of more than one enterprise. The software configuration pursuant to the requirements of a specific enterprise and the respective accountant starts with company information, address, identification code, manager, branches. The next step comprises the choice of chart of accounts and configuration of synthetic and analytical accounts. This is followed by the entry of initial balances of the accounts. The software allows entry of detailed information for the enterprise's accounts – receivables and payables by partners and by documents.

The primary accounting information is entered through a multifunctional counting (accounting) window allowing simultaneous accounting entries and tracing the accounts' balances. The option for automatic cash payment of receivables and payables saves time and ensures information for correct preparation of the statement of cash flows. The developed auxiliary panel provides the user timely information about any errors omitted and hidden functionalities. There is an option to create the so called "templates" – enterprise's business-specific accounting entries that are made on daily basis and repeatedly. Thus, for example, the enterprises operating in the field of trade usually have to make accounting entries relevant to the purchase and sale of goods; the transport companies – the purchase of fuel and lubricants and posting the revenue from transport services provided. Regardless of their scope of business, all enterprises charge employee benefits, depreciation expenses, etc. on regular basis, and it is also quite appropriate to create templates for these types of expenses.

The software ensures unlimited opportunities to edit data thus implementing overall monitoring on information entry. The option to arrange already edited or deleted accounting operations chronologically (by date) or by their unique number is another advantage of **Microinvest Delta Pro**.

The currency accounts and revaluation module so developed facilitates accounting operations thus allowing automatic revaluation of receivables, payables and cash denominated in foreign currency pursuant to the requirements of both national and international accounting standards.

Further to currency revaluations, **Microinvest Delta Pro** also performs automatic accounting operations for:

- Settlement of receivables and payables by partners and by documents, in national and in foreign currency, in cash or by bank;
- Monthly closing of accounts for reporting calculations under the Law of the Value Added Tax and calculating the results for the period – tax to be paid or tax to be refunded;
- Closing of accounts for reporting of expenses by segments and of operating revenue;

The accountant's operations are further facilitated by the modules developed to check-up missing documents by numbers or to check-up doubled documents, as well as the options toe check the validity of a contractor's identification code, VAT number or PIN. The option allowing import of accounting operations from other software products is also very useful, since this saves a lot of time to be spent by the operating accountant to manually enter the primary information.

With view of the main and auxiliary registers, **Microinvest Delta Pro** allows generating a chronological register of all economic operations performed and by preset

criteria – by specific account, document number, date or period, partner or type of transaction. An "Analytical Register" module is developed, which generates systematic registers for the movement and balances within the accounts.

On the basis of data entered from the primary accounting documents it is possible to automatically generate a major book and a trial balance for a period selected by the user, which contains only synthetic accounts or synthetic and analytical accounts.

The module for automatic generation of VAT schedule-return, log of purchases, log of sales and VIES return allows the enterprises having VAT registration, to prepare the registers required by the Law of Value Added Tax immediately after the entry of information about any purchases and sales performed. The user has the option to choose a directory where to save the VAT registers, which facilitates file finding and emailing.

The final stage in each accountant's work – the preparation of financial statement is also facilitated and to a great extent automated. Through the in-built formulas, accounting information is summarized in compliance with the applicable accounting standards, the requirements of the National Institute of Statistics and the National Revenue Agency. This module is open and allows the accountant to determine what information is to be summarized in each cell of the statements.

Auxiliary registers supported by Microinvest Delta Pro:

- **Fixed assets –** inventory cards of amortizable fixed assets are created where information about assets is entered – name, inventory number, location, accountable person, date of acquisition, amortized value, depreciation method, accounting and tax depreciation standard and quota. There is an option for automatic charge of depreciation for entered fixed assets and for automatic retrieval of inventory book, accounting and tax depreciation scheme, schedule of monthly depreciation quotas – both accounting and tax;
- **Inventories –** information about the movement of enterprise's inventories is originally entered in the software from the "counting" window, and the summarized information about suppliers, sales, ordering and inventories on hand can be retrieved from the Schedules menu;
- **Cash register –** this module provides information about cash on hand in the enterprise's cash register. Primary information about any cash operations is also entered in the "counting" window. The schedule developed within the software, called "cash book" presents summary of cash movement and balances.
- **Offsets –** this auxiliary register allows the accountant to prepare schedules for receivables and payables by selected partner, document number, at specific time or for a specific period.

The reporting process technology, the ways for entry of accounting information, the existing functionalities and options for automated information processing, as well as the generation of flexible, customer-friendly main and auxiliary registers featured by **Microinvest Delta Pro** are presented in the table below:

Table 1. Microinvest Delta Pro Form of Accounting – technology of reporting process

1st STAGE	**2nd STAGE**		**3rd STAGE**
Input Primary Information	**Process Information Processing**		**Output Summarized Information**
	MAIN REGISTERS		
Counting (accounting)	Chronological register		Financial statements
Accounting correspondences (templates)	Analytical register		Statements to the NIS
Counting editing	Main book		Return under the Law of VAT, register of purchases, register of sales, VIES return
Counting rearrangement – by number, by date	Trial balance		
	AUXILIARY REGISTERS		
Automatic operations – payments, closing of VAT accounts, currency revaluation, closing of income and expense accounts	**Fixed assets**	I nventory cards	Inventory book, SAP, DAP
Check-up of doubled, missing documents	**Inventories**	c ounting	Schedules of supplies, sales, ordering (methods) and inventories on hand (balances)
Automatic import of data for contractors upon entry of EIK and active access to the Internet	**Cash register**	c ounting	PKO, RKO, Cash book
Check-up of validity of VAT number, EIK, PIN, bank account number	**Offsets**	c ounting	Schedule of receivables/ payables by partners and by documents
		p ayments	
Import of operations from text file and MS Excel	Export of accounting registers in the following formats: *.rtf, *.xls, *.xlsx, *.pdf, *.xps, *.html		

2.3 Benefits of Microinvest Delta Pro Accounting Software

The major benefits of **Microinvest Delta Pro** are its unique characteristics distinguishing the software form the remaining accounting products and outlining its leading positions in the following directions:

- It operates with SQL server thus allowing to work with a common database (DB) from remote work places;
- It features user-friendly and simplified intuitive interface;
- It allows dynamic revision of fixed assets tax and accounting depreciation scheme; it allows adjustment of specific monthly quotas and conversion of the scheme according to the changes so made;
- All processes of the software are based on the basis of the chart of accounts thus preserving the accounting concept of programming;
- The counting window is a key module for all functions of the software thus avoiding complex settings before start working with the software;
- Online support is available through the in-built feedback form;
- There is an option available for additional processing of standard accounting schedules through filters and screening by components and analytical levels.

Conclusions

This report emphasizes on the concept and directions within the organization of accounting operations. The capacities, functionalities and benefits of the **Microinvest Delta Pro** accounting software as presented above are an inherent technological challenge faced by small and medium size enterprises when organizing their reporting process. **Microinvest Delta Pro** ensures each user quick adaptation and orientation, entry and exchange of large information arrays within short terms, flexibility of data editing, as well as quick and prompt generation of schedules and registers for the needs of the enterprise's management.

References

1. Belmer, F., Common Theory of Accounting, Third edition, Varna, 1932, page 587-590.
2. Trifonov, T., Accounting Theory: Philosophic and Economic Principles and Modelling, University Publishing House Stopanstvo, Sofia, 2008, page 13.
3. Microinvest LTD, https://www.microinvest.net

Business Downtime and its Impact on Business Organizations

Asen Bozhikov

Abstract. Modern business organizations are driven by market forces and the customers. To survive and get advantages in this complex business environment they need to be competitive and to deliver services with efficiency and swiftness. That's why every business downtime due to unexpected or unplanned events means possible loss for the organization which directly affects the revenues and cash flow. Calculating the financial impact from a possible business disruption, which leads to downtime, could help the organization in finding alternative ways to react and recover its services as fast as possible. The aim of this paper is to provide an overview of the possible tangible and intangible costs which are part of the total costs of downtime. Dealing with downtime could be depicted in the business continuity/disaster recovery plan of the organization, if there is one in place. Setting an acceptable time span for customer services to be up and running again after downtime is important part of the plan.

Keywords: downtime, disaster recovery, business continuity.

1 Introduction

Over the last 20 years businesses of all sizes are getting more dependent on their fast-growing IT infrastructures. Whether it's a web site, reservation system, billing system, CRM, ERP, SCM or other form of advanced automation, business organizations' success is extremely linked with the availability of those systems, services and data, connected with them. Today the abbreviation 24x7 or even 24x7x365 which means continuous operations of an IT system or an e-commerce web site is seen quite often. Unfortunately IT systems are susceptible to failures. A disruption in any part of the organization's IT infrastructure leads to downtime which could harm the business in many aspects.

There are many events which could cause different types of failures in the IT infrastructure of a business organization. Most executives focus their attention on natural disasters like floods, hurricanes, tornados, etc. Experienced IT managers know that there are other disruptive events like power outages, terrorism, employee sabotage, data fraud, failure in disk drive or network components, etc. The business implications of these disruptive events, often called disasters, are clearly depicted with the term downtime. It is important for the business organizations to analyze every possible downtime thread, mark it as internal or external for the organization and prepare to deal with it in a reasonable timeframe.

Every organization should have a plan what to do if any of the above mentioned scenarios occurs and how to recover its services and data as fast as possible so that the downtime is minimized in a reasonable timeframe. In other words the organization should have business continuity/disaster recovery plan which is a step by step guidance to successful recovery and downtime mitigation in the case of a disruptive event.

2 Downtime

The term downtime is used to describe the total time between the occurrence of a failure and the repaired item being put back into operation. Downtime is divided in two groups: planned and unplanned (see Fig. 1). The planned downtime includes normal IT infrastructure operations like backup and security; maintenance of the software, hardware and network components; and some periodic activities like hardware or software upgrades which usually could take longer time opposed to the previous two. The planned downtime is made according to a schedule on a regular basis. It's done in the appropriate timeframe, usually when the system is idle during the night. The duration of this type of downtime is typically estimated and known. The unplanned downtime, as the name implies, appears when unforeseen circumstances lead to loss of connectivity to the system or loss of functionality of the system or other things that disrupt normal business operations of the organization. The duration of the unplanned downtime is rarely known in advance. That's why it is a serious threat to the normal execution of the organization's IT operations.

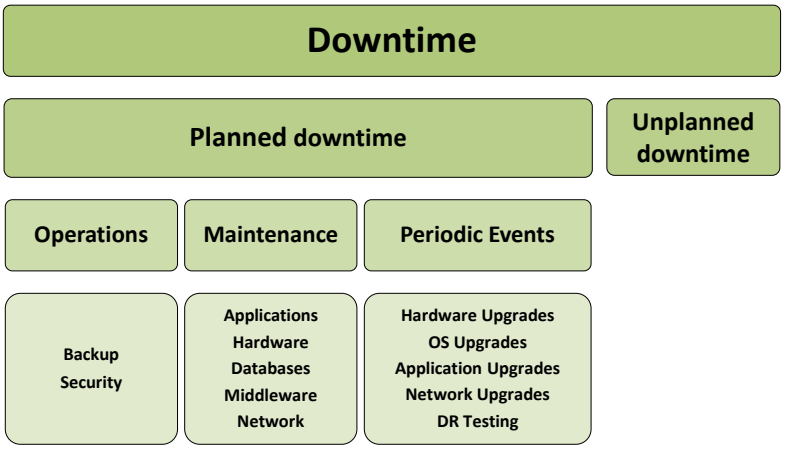


Fig. 1. Planned and unplanned downtime.

The unplanned downtime accounts for 10% to 15% of all downtime (almost 50% of the unplanned downtime is caused by a human error). Nevertheless its unpredicted nature could case enormous damage for the organization. That's why calculating the cost of downtime should be a critical task for every business organization.

According to Gartner's survey through 2015, 80% of downtime impacting mission-critical application and services will be caused by people and process issues, and another 50% will be caused by change/configuration/release integration and hand-off issues. This will be the result from the growing adoption of virtualized solutions and cloud-based services in business organizations.

3 The impact of downtime

Downtime has different levels of impact on business organizations. The impact will depend on the size of the organization and the industry in which it operates. Clearly, a small-sized business organization wouldn't suffer serious losses if its IT infrastructure is down for an hour but that would be disruptive for a big corporation. Usually, downtime is linked only with loss of revenues and customers but there are other key indicators that are impacted by the downtime (mostly unplanned). Thorough understanding of the impact of downtime includes calculating the possible costs of downtime.

Calculating the real costs of downtime is not an easy task. There are many costs that should be considered, some of them really hard to assess. Downtime costs are classified in the following two groups:

- Direct (tangible) costs – some of them are:
  - ✓ Lost revenue
  - ✓ Lost facilities
  - ✓ Lost wages
  - ✓ Legal penalties linked with the service level agreement
  - ✓ Marketing costs
  - ✓ Cost of recovery etc.
- Indirect (intangible) costs – some of them are:
  - ✓ Lost business opportunities
  - ✓ Loss of employee morale
  - ✓ Decrease in stock value
  - ✓ Loss of customer goodwill
  - ✓ Brand damage etc.

There is an equation proposed to directly calculate the costs of downtime. It is not so accurate though because it uses a constant value for the costs of lost revenues, productivity and reputation. The equation is as follows:

$$Costs\ of\ downtime = ((SH + WH) \times H) \times 2) . \quad (3)$$

Where:

SH = average sales lost during an hour of system downtime

WH = total hourly wage of all idle employees during an hour of system downtime

H = estimated number of hours of system downtime during a year

2 = take into account the cost of lost business, lost productivity and damaged reputation.

Another important factor to consider is the labor productivity. If an application or a system is offline and is not working after a downtime scenario this means that the employees who use the system can't execute their everyday tasks. They're idle and their productivity is reduced to zero but they continue to receive full pay. Meanwhile other employees who are not affected by that system's downtime will continue their normal work. A simple formula to calculate the average labor cost of downtime is:

$$Labor\ cost = P \times PA \times EC \times H . \quad (2)$$

Where:

P = number of people affected
PA = average percentage they are affected
EC = average employee cost per hour
H = numbers of hours of downtime.

This equation needs to be calculated for every department in the organization which will guarantee higher degree of accuracy.

By using another equation it is possible to calculate the potential annual revenue loss due to downtime:

$$Lost\ revenue = (GR/TH)\ x\ I\ x\ H\ . \qquad \textbf{3})$$

Where:
GR = gross yearly revenue
TH = total yearly business hours
I = percentage impact
H = numbers of hours of downtime.

Some organizations provide their customers with a service level agreement (SLA). In general the SLA specifies the necessary course of action to be taken by the parties in case the terms are not met by either party. More specifically, the SLA should define the minimum acceptable level of service availability. If a disaster strikes and the organization can't provide that service to the customer because of a longer downtime then it could be a subject to some legal penalties. That's why the organization needs to take into account the amount of money paid for penalties.

The availability of a system or service could be calculated with the following equation:

$$Availability = \frac{Uptime}{Uptime\ +\ Downtime}\ . \qquad \textbf{4})$$

The uptime is the total time in hours when the system is up and running (active). The downtime here includes the total time when the system is not working either because of planned or unplanned downtime. The above equation shows quite obviously why the SLA is connected with the downtime.

These days, customers usually insist upon 99.99% of availability for a service which forces business organizations to analyze every possible downtime thread and implement high availability. Table 1 shows some common values for availability and the corresponding values for the possible duration of the downtime (both planned and unplanned).

Table 1. The percentage of availability and their time equivalents.

Acceptable uptime (%)	Downtime per day	Downtime per year
95%	72.00 minutes	18.26 days
99%	14.40 minutes	3.65 days
99.9%	86.40 seconds	8.77 hours
99.99%	8.64 seconds	52.60 minutes
99.999%	0.86 seconds	5.26 minutes

Indirect costs are hard to evaluate as they include the loss of potential customers, damaged reputation and employees' morale. The appearance of unplanned downtime in a business organization, especially if the downtime is reported in the medias, could also reflect on a company's stock price. A classic example of this scenario is Amazon's two hours of downtime in 2008 which led to the fall of the company's share price by 4.1% on the same day. The same happened with eBay in 2001 and happens with many other companies worldwide nowadays (Apache Corp. in 2011).

Damaged reputation is a result of the fact that customers are much more likely to talk about bad customer experience than good. That could have huge impact on the organization's reputation and could drastically decrease or even stop the flow of new potential customers of the organization's products and services. It could take months, even years for the business organization to reverse that negative effect. It is often a good idea if the organization which has suffered unplanned downtime tries to invest some money for research and surveys and investigate the long-term effect of the damaged reputation. The costs for conducting a research or a survey will be a result of the downtime, that's why they need to be included in the total costs of downtime.

Another indicator that is difficult to measure is the employees' morale. If the data or the systems (or any other component of the IT infrastructure) which they use are not reliable or unavailable most of the time, that could damage their morale. As a consequence of that they might decide to leave the organization and find a new job in another business organization. Although there is no obvious cost in that situation a thorough review will depict that the loss of an employee could result in lost output until a replacement is found and also the organization should spend money to hire and train the new candidate.

According to a survey of CA Technologies among 200 organizations from all sizes and industry sectors across North America and Europe:

- Organizations are collectively losing more than 127 million person-hours each year through IT downtime and data recovery (545 person-hours on average);
- Each business suffers as average of 14 hours of downtime per year, during which time employees are only able to work at 63% of their usual productivity;
- During these periods, when business critical systems are interrupted, companies estimate that their ability to generate revenue is reduced by a third (32%);
- In this post-outage period when data recovery is taking place, company revenue generation is still severely hampered, down by an average of 25%;
- 50% of organizations revealed that IT outages can damage the company's reputation while 18% thought this would be "very damaging" (see Fig. 2);

- 44% of respondents believe IT downtime can damage staff morale and 35% report it can harm customer loyalty (see Fig. 2);
- 87% of businesses indicated that failure to recover data would be damaging to the business. 23% said this would be "disastrous";
- Organizations in the public sector experience the longest downtime (18 hours) and recovery time (11 hours) per year. The retail, finance and manufacturing sectors have shorter periods of downtime and recovery time – 22, 21 and 20 (total) hours respectively.

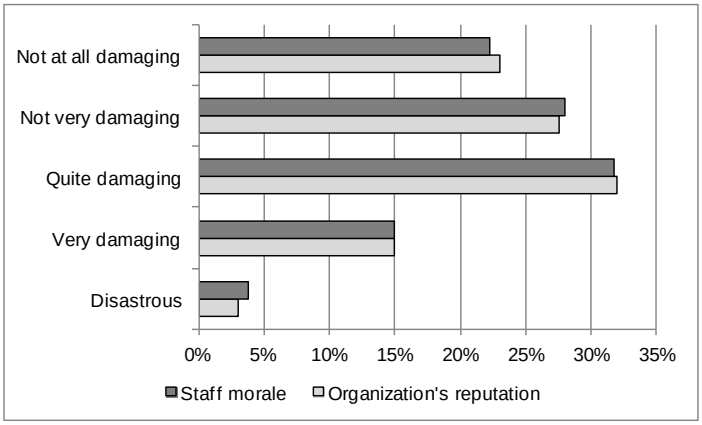


Fig. 2. Downtime effect on the organization's reputation and the staff morale.

The results of the above research show that the downtime is a real threat and it happens no matter the location of the business organization. The research proves that the impact of downtime is beyond the loss of revenues and customers, but it also affects employees' productivity and morale, the organization's reputation, etc. That's why business organizations worldwide should be prepared for every possible disruptive event that could lead to downtime. Of course 100% of threats coverage and service availability is hard to achieve but being close to that – like 99.99% or 99.999% should be a priority goal for the executives of the organization.

Again, it is important to underline that the availability percentage (possible downtime respectively) is connected to the size of the organization and the industry in which it operates. In general, small and medium sized organizations can't afford the expensive solutions for high availability which large organizations have adopted. The growing implication of virtualization and cloud services on the IT infrastructure of the smaller organizations reduces that gap. Increasingly nowadays it is not the actual size of business organizations that defines how important their data and systems are, but the line of business they operate in. Many smaller organizations rely on IT systems to generate revenue, and the outage is proportionally just as great as in a large multi-national company.

In conclusion downtime can appear unexpectedly at any time and in any form and can disrupt normal business operations in the organization. A possible downtime affects not only the revenue but also labor productivity, employees' morale and organization's reputation. When calculating the costs of downtime all of these tangible and intangible costs should be put together in the equation. No matter the downtime scenario business organizations must plan for such contingencies and what they will do about them when they emerge.

References

1. How downtime effects your reputation, http://esellermedia.com/2012/09/12/how-downtime-effects-your-reputation/
2. Maintaining High Availability, http://msdn.microsoft.com/en-us/library/ee799074(v=cs.20).aspx
3. Mirsha R.: Reliability And Maintenance Engineering. New Age International Publishers, New Delhi (2006)
4. Oggerino, C.: High Availability Network Fundamentals. Cisco Press, Indianapolis (2001)
5. Schmidt, K.: High Availability and Disaster Recovery: Concepts, Design, Implementation. Springer, Berlin, (2007)
6. Sople, V.: Business Process Outsourcing: A Supply Chain of Expertises. Phi, New Delphi (2009)
7. The Avoidable Cost of Downtime, http://www.ca.com/~/media/Files/SupportingPieces/acd_report_110110.ashx
8. The Avoidable Cost of Downtime, Phase 2, http://m.softchoice.com/files/pdf/brands/ca/ACOD_REPORT.pdf
9. Understanding Downtime, http://www.mik3.gr/docs/VWP_Downtime.pdf

Developing Marketing Strategies with a Marketing Simulation Tool in a Master's IT Class

Vanya Slantcheva-Baneva

MT&M College
126, Tzar Boris III Blvd., 1612 Sofia
vanya.slantcheva@gmail.com

Abstract. This report examines the outcomes of running a specialized marketing simulation tool in the classes of Postgraduate IT students for a period of three consecutive academic years. A segmentation of the developed marketing strategies is provided in order to consider the performance mode of IT students in playing the simulation and the learning objectives they pursue by means of 'concrete experience' paradigm. A qualitative assessment has been made and the outcomes convincingly reveal that students do not "think about the whole picture" of building up a marketing strategy and respectively execute it in a short- or medium-term tactical pace. IT students begin playing the simulation with a strategic view often maintaining the status-quo position of the company under consideration. Their playing decisions are primarily based on checking out the dashboards of profitability and customer satisfaction in each quarter that notably reflects the direction – up and down – of their inputs. That feedback has made them encouraged or discouraged to continue, rather than to test the marketing strategy in an integrated manner.

Keywords: business-to-business marketing strategy, marketing simulation, postgraduate IT students, experiential learning.

Introduction

Learning to develop marketing strategies in a disciplinary context is a demanding educational process. Besides the use of conventional teaching approaches, marketing strategizing becomes an object of non-traditional interventions [4], such as Marketing simulations and Business games. The evolution from a conventional lecture-based course to an interactive, computer-based mediation entails an adoption of 'concrete experience' and 'experimental learning' categories in higher education.

The report considers selected elaborations that represent critical outcomes by means of using interactive tools into the educational process. Tonks's reveals the 'experimental learning' category as the non-conventional modus of students' own experimentation [5] when exercising with Marketing simulations. In addition, Lin and Tu propone the 'autonomous, visibly thought learning" through games as an interactive way to deliver knowledge to students [4]. The real measures of claiming the mediation of the interactive tools in class are pursued learning objectives.

Introducing the online-based interactive tool *Marketing Simulation: Managing Segments and Customers© (MS©)* into classes of postgraduate IT students emerges as a valuable academic experience of implementing the 'experiential learning' paradigm. The simulation

was conducted during three consecutive academic years – 2009/10, 2010/11 and 2011/12 and throughout that medium-term period IT students participated in the non-conventional mode of developing marketing strategies.

A need for segmenting marketing strategies postgraduate IT students developed when simulating with the MS© has been identified in order to draw attention to the issue of students' perception of classroom sessions and debrief support, and of student attitudes to playing the simulation itself. The segmentation is based on the standard typification elaborated by Das Narayandas [1, 2]. Further, a research on the nature of the learning objectives that the postgraduate IT students are to pursue is explicitly provided. On the one hand, to learn about the application of Marketing concepts and techniques in an integrated manner, while operating in an unpredictable and competitive environment, requires the development and implementation of a data-driven analysis and planning system. On the other hand, to understand and practise personal skills in marketing management via individual (or team) work in the simulated environment requires coping with the problems of uncertainty and self-debriefing [6]. To affirm which learning objectives have been pursued in playing the MS©, a parallel is to be drawn between the students' involvement by means of their enrolment, participation and activity, on the one hand, and their personal scores, on the other.

1 Marketing Simulations in the Core of a 'Concrete Experience' and 'Autonomous Learning' Paradigm

1.1 Marketing simulations and the category 'experiential learning'

Marketing simulations are an interactive, computer-based model of a marketplace in which students analyze available data, take decisions concerning marketing variables, and then receive results on performance. The model is basically a series of mathematical formulae, which represent the structures, relationships and sensitivities in the simulated environment. The model can become a "black box" which the students are seeking to unravel [5].

Marketing simulation exercises embed ideologies that refer directly or indirectly to learning objectives. Learning objectives represent a real outcome measure of an experiential learning. Approaching the 'experiential learning', Tonks stresses the importance of immersion in *a concrete experience*. Observation and reflection on such experience leads the participants to the creation of their own active experimentation [5] and autonomous learning [4].

'Participants' are the recipients of a learning experience – commonly termed 'students'. Experiential learning is considered, employed and evaluated in relation to the students, those on the receiving end, but the 'participants' also include those who deliver – the instructors [5].

'Experiential learning' has been noted for its relatively high level of involvement by participants, rather than the more passive style that is a feature of more conventional learning experiences. By their very nature, marketing simulations exercises that are interactive and dynamic, and which require students to take responsibility for outcomes, tend to fall into the category 'experiential learning'.

From that perspective, Tonks traces and explains *his* understanding of marketing simulations into seven phases. Experiences could be best translated as a desire *to transmit*

basic core knowledge [5] and to provide awareness of the interactions between decision variables.

Although the requirements of 'working in groups', 'coping with uncertainty' and the 'centrality of planning' exist independently of an educational rationale, *a professional understanding is demanded* [5]. Therefore the subject matter of a Marketing course, supported by conducting marketing simulation, should have an instrumental, managerial, pragmatic and vocational orientation.

Marketing simulation exercises are worthy in *a pragmatic sense* [5], as they approximate to working reality. The responsibility to be aware of the realities and needs of the world of work are to be manifested in stated learning objectives [5].

The *process and context* [5] of marketing simulations makes the issue explicit. Learning objectives of a course design produce a significant shift: from "developing and implementing a data-driven analysis and planning system" towards "coping with the problems of uncertainty and managing the group involved effectively and efficiently".

The considered reflection of the process and context in conventional learning (e.g. case studies) rises to the surface *critical skills* [5]. Although the prevailing part of the students respond very positively to the engaging, entertaining, fun aspects of the 'concrete experience' when running the marketing simulation, for some students the surface approach is taken to 'deep-level' learning [5].

1.2 Business Simulation Games and Game-Based Learning

In addition to marketing simulations, Business Simulation Games (BSGs) also present an effective alternative to traditional teaching methods. They provide a link between abstract concepts and real world problems. They offer practicality to the learning setting and give students the chance to practice decision-making cases. In their study, Lin and Tu pointed out that most students thought that they had gained a lot of management knowledge. According to the authors, learning through games becomes an interactive way to deliver knowledge, and game pathway would provide four advantages to support learning: (a) making knowledge accessible; (b) making thinking visible; (c) making learning fun; and (d) promoting autonomous learning [4].

The researchers Lin & Tu applied a soft-laddering technique to analyze how students perceive the outcomes of BSGs use. In the laddering technique, respondents are delineated by their behaviours as 'gamers'. Students have been asked to differentiate game attributes in terms of their specific and individual importance, being questioned "why is this important to you?" The expressed goal is to determine sets of linkages between key conceptual elements across the range of 1) attributes, 2) consequences, and 3) values for the students. According to the frequency, i.e. the average number of times a ladder was mentioned, authors point major linkage paths out. As an important attribute when using BSGs, students outline 'Teamwork', which reflects in the 'Emotional Exchange' consequence and brings about the value of 'Interpersonal Relationships'. In the same line, the decision-making process of considering 'Market Diversity' has as a consequence 'Multi Thinking' disposition and benefits the value of individual 'Sense of Accomplishment'. Respectively, the linkage between the attribute, 'Simulated Business Operations', the consequence – 'Understanding Business Concepts', and the value of 'Fun and Enjoyment of Life' reveals that students not

only can learn or combine knowledge in the process of BSG education, but also feel the ultimate value of fun and enjoyment of life [4].

1.3 The Marketing Simulation: Managing Segments and Customers©

The Marketing Simulation© is an online-based interactive tool in which students play the role of the CEO of a manufacturer of medical motors Minnesota Micromotors [1]. Throughout the simulation, students make decisions regarding MM's marketing and operations strategy. They make important decisions – regarding product design, pricing, discount structure, marketing expenditures, sales-force size – that should collectively support an overall marketing strategy designed to achieve a combination of sustainable revenues and profits over 12 fiscal quarters. Performance is measured using both qualitative and quantitative criteria, including profitability, revenues, unit sales, market share, and customer satisfaction [2].

In this simulation, students define and execute a business-to-business marketing strategy. Customers are divided into market segments based on their requirements for two key motor performance features and price. Students must analyze each market segment and decide which new customers they want to acquire while also considering the loyal customers they must retain [1]. A successful go-to-market strategy requires careful consideration of a variety of interdependent factors. Students set a list price and then set discounts for each large market segment and for a segment of small customers who purchase through distributors. Students allocate sales and marketing resources for each targeted segment including setting the level of spending on marketing communications and market research. Students can listen to customer feedback through dynamic video interviews and gain important insights into the effectiveness of their marketing strategies. Ultimately students must achieve a sustainable revenue stream to maximize cumulative profit for the company [1].

The simulation is designed as a single-player exercise although the instructor can have the students play it in teams. Throughout its duration, the simulation provides students with an array of tactical decisions to make and retain a strong focus on marketing practice and theory. The structure of the simulation is interactive and enables multiple opportunities for self-debriefing and learning. The simulation is also flexible enough to allow students to use a variety of strategies successfully. There is no single correct solution [2].

Table 1. Learning objectives of the Marketing Simulation that vary according to the instructor's goals and the nature of the class [2].

No	Learning objectives
1	Understanding the link between marketing strategy formulation and effective implementation and execution
2	Using segment / customer needs analysis to make product design decisions and associated trade-offs
3	Understanding segmentation, targeting and positioning
4	Pricing strategies – setting and changing price
5	Managing channel conflict and maintaining consistency across multiple go-to-market channels
6	Assessing the importance of marketing strategy to match changes in the market over time (a static view is not sustainable)
7	Using hard and soft metrics to measure firm performance
8	Responding to competitive actions / reactions

9 Trade-offs between customer acquisitions and retention
10 Distinguishing the differential impact of lumpy, fixed and variable costs in sales, in marketing, and in production
11 Appreciating the relationships among customer satisfaction, customer buying patterns, customer loyalty, and firm profitability

2 Developing Marketing Strategies using Marketing Simulation©

The need for segmenting student marketing strategies has been recognized in order to disclose the students' perception of the classroom debrief support and mode of their further proceedings. In that respect, a question is raised whether the postgraduate IT students prefer to play the simulation pragmatically or holistically. Running pragmatically the MS© is considered to be a subordination of an individual performance. The holistic *juxtaposition* is used to display students' ability to run the simulation having an integrated mind-map of developing and executing a marketing strategy over three fiscal years. It means considering the scope of the simulation by managing segment and customer behavior; monitoring competitive behavior; understanding market and considering market conditions; using marketing research; making investment decisions; leveraging by pricing decisions; and managing channel conflict.

2.1 Enrollment of Students

Launched by Harvard Business Publishing for Educators in 2009, the *Marketing Simulation: Managing Segments and Customers*© [1] has been conducted in the classes of postgraduate IT students during three consecutive academic years – 2009/10, 2010/11 and 2011/12. The data collected within that medium-term period of observing a similar process of 'experiential learning' in a Marketing Management course need to be analyzed and reported.

The IT students who participated in the MS© were doing their Master's in "Technological Entrepreneurship and IT Innovation", "E-business and E-governance" and "Information Systems". It is relevant to note their majors because the dominant number of students taking part in the Marketing Simulation are assigned the task of developing start-up business plans and go-to-market strategies. That educational background *should* get them motivated for the experiential learning through marketing simulation tools. However, the MS© is optional and paid, and therefore students are more likely not to be enthusiastic, but rather unwilling due to pragmatic considerations, such as not getting overloaded with academic duties. Nevertheless, the price is reasonable – $12.50 for six-month full access with a lead option for more than half of any class. The actual number of participants might reach the moiety. But there are objective and subjective factors that influence the student's vote. And again, IT students are inclined to pay for the simulation, but not necessarily to play it, as it is shown in Fig. 1. Hence, the instructor should be concerned when students pay for their enrollment but actually run the simulation apathetically, moreover when they do not run it at all.

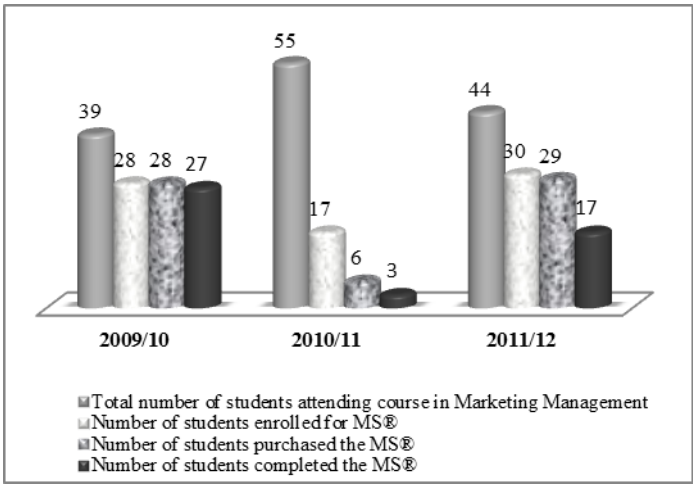


Fig.1. Overall student intentions to participate in playing online-based Marketing Simulation during the three sequential academic years between 2009 and 2012.

The introduction of the Marketing Simulation© in 2009/10 pulls 72% of students to purchase it. Although it appeared to be in demand, the purchases of the simulation dramatically dropped to 11% in the next 2010/11. It is more likely to consider it as a precedent because in 2011/12 the number of students purchased the simulation raised to 61%. However, more than a five-year data record is required in order to outline a trend of the actual postgraduate IT students' interest in learning by means of the marketing simulation tool.

The MS© exercise is optional, and the figures above could be considered as decent. Doubtlessly, should students consider their extracurricular tasks, they decide whether to purchase and play the simulation. Although they were not post-interviewed about the intentions and biases that had influenced their decision-making, IT students demonstrated them convincingly in an informal way. Basically, they pay attention to objective factors, such as attendance, employment, participation in student contests, etc. Certainly, students are also concerned by some subjective factors, such as motivation, interest in the course, individual perception of the instructor's capabilities and charisma, family.

Registration and purchasing the MS© does not necessarily mean that students need to learn by it. An important evidence in this respect is the number of students who actually ran and had the simulation completed. In the academic year 2009/10, student activity reached about 100 per cent – 27 out of 28 enrolled students, ran and completed the simulation. In 2011/12, student activity, compared to 2009/10, declined almost twice: the number of students who purchased the simulation was 29, but only 17 of them completed it. The academic 2010/11 is a precedent, as mentioned above: the number of students in class was the highest one – 55, but the number of those who purchased the simulation, was very modest: 6; and hardly 3 of them played and completed it. The whole picture is delineated in Fig. 2.

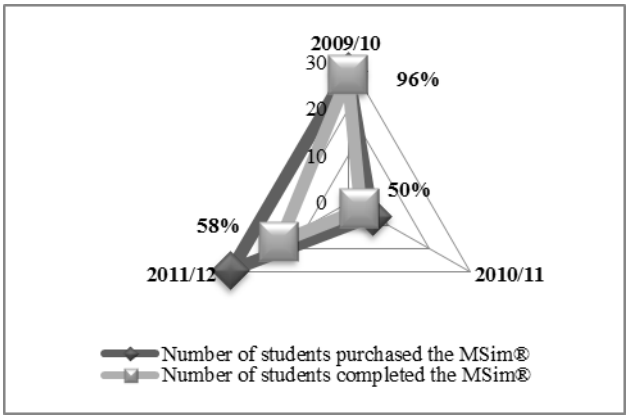


Fig. 2. Relative share of students who actually completed the Marketing Simulation© after purchasing it, compared to three sequential academic years between 2009 and 2012.

As it is displayed on Fig. 2, there is an obvious challenging discrepancy between the "black" triangle that covers the number of IT students who purchased the MS© and the "white" one – with those who completed it. The number of "whites" is growing smaller than the number of 'blacks'. Approximately one half of the students do not complete the MS©, despite of having purchased it. This behavior is repeated in two consecutive academic years: 2010/11 and 2011/12.

Zooming in the 2011/12 corner of the triangle, figures claim that 17 out of 29 students have completed MS©, or alternatively, above 40% of the students have not completed the simulation. Four students have not run the MS© at all, although they have purchased it. The highest number of rounds (running MS© through 12 fiscal quarters) is 41 rounds per student. The lowest one is just once. Hence, the average number of rounds per student is 8. The most effective student runs totally 7 rounds and reaches the highest score in the group twice (81 point). The most diligent student runs the MS© 41 rounds and reaches the score of 80 points. Hence, the average efficiency of the rounds – in 8-9 completed rounds the individual result is above 50 points. And the overall group activity hesitantly surpasses 50 per cent.

The point raised here is that Master's IT students ultimately refuse to run the MS©. Hypothetically, their arguments could be: ill-instruction, loss of interest, feeling of being not involved enough; MS© is a demanding and time-consuming process of learning. The shift away from hands-on, competitive, engaging activity towards more considered reflection on process [5], e.g. case studies, did not meet with universal approval. The previous experiences of some students, their learning styles and their resulting expectations are such that more conventional approaches create less stress [5]. In this respect, another research is required to differentiate the threats of using the MS© as an integrated instrument of an experiential learning in a course in Marketing Management.

2.2 Segmentation of Marketing Strategies adopted by the Postgraduate IT Students Exercising with the Marketing Simulation©

There are three general business-to-business marketing strategies that the students are debriefed to follow: 1) 'Status quo Strategy', 2) 'Increase Price Strategy', and 3) 'Commodity Play Strategy' [see Fig. 3]. Resolutely, Master's IT students from the three academic classes adopt 'Status quo Strategy'. It is the strategy through which students understand what the company 'has in store" so as to be long-lasting. The students primarily focus on managing the Large-Volume Customers (LVC) in the face of company's Loyals and in accordance to the attractiveness of the other existing market segments. While strategizing and executing the *status quo*, students intuitively bring about the LVC Retention Spending as it is at the lowest risk and belongs to an appropriate cost effective structure. They adopt the 'Status quo Strategy' for a medium-term period of three fiscal years (the length of the simulation), but they understand they should consider that market situation and customer demands change over time and adequate decisions with specific product features improvements, investments, price, and sales-force accents are necessitated.

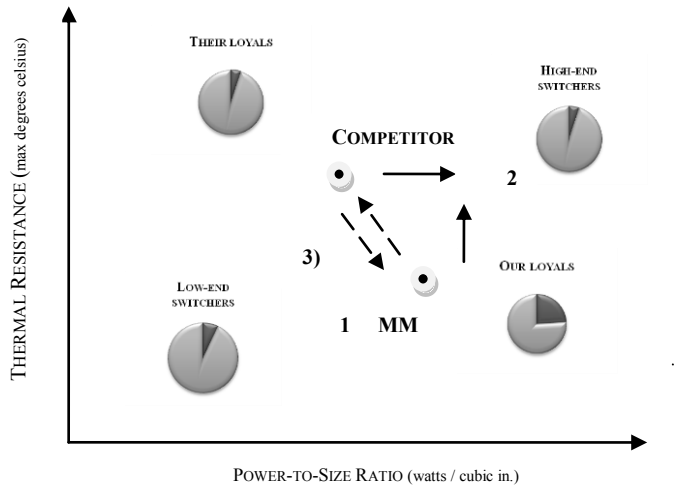


Fig. 3. General business-to-business Marketing strategies that students execute using the Marketing Simulation Tool. The scheme is developed by Das Narayandas [2].

Moreover, a pretty inherent manner to maintain status quo yet remains to set up ambitious goals as to expand company's total market share at a high profit. Gaining a portion of a market share means, at least, obtaining new customers among the pursued – new or existing – segments, and consequential product investment. "Up to 12% market share" results display defensive and retentive decision on the part of the students. Individual best scores achieved endorse properly applied tactics of "retention of existing customer" and "market penetration" strategy. As an evidence, 70 % of the 2011/12 class hit $50,000,000 – $69,999,999 cumulative revenue.

The outcomes convincingly reveal that students do not "think about the whole picture" of building up a marketing strategy, and respectively implement it in a short- or medium-term

tactical pace. Definitely, IT students begin playing the simulation with a strategic view maintaining the status-quo position of the company. Their playing decisions are primarily based on checking the dashboards of profitability and customer satisfaction in each quarter that notably reflects the direction – up and down – of their inputs. That feedback has made them encouraged or discouraged to continue, rather than to test the marketing strategy in an integrated manner.

Despite of that students adopt one and the same "Status Quo Strategy", they have differently distributed individual scores. There is no single correct solution. Students do not base the Investment policy in product performance on high levels of cumulative profit. They operationalize the improvements in product features. The IT students are focused on serving and paying for customer satisfaction, grounded on short-term profits. In the core of executing "Status Quo Strategy" is pricing, product investments and sales-force management considered entirely from the perspective of serving Loyals and the mixed-bag customers [3]. Alternatively, a transfusion towards 'Price Increase Strategy' appeared to be beneficial if high-profit customers are the targets.

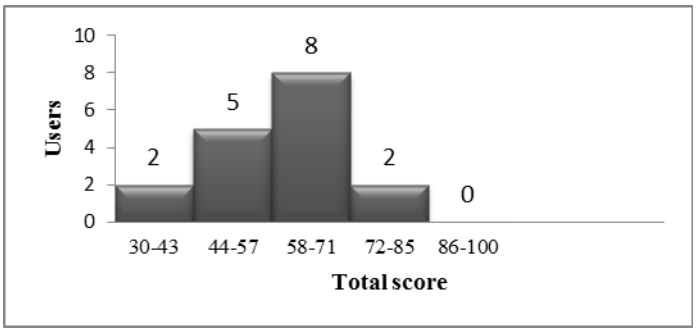


Fig. 4. Distribution of class results by best individual score (2011/12).

The prevailing and active part of the 2011/12 class (47%) has got above-average individual scores – in the quartile of 58-71 points [see Fig. 4]. There is no excellent score – above 85 points. But there are two interesting playing behaviours with the highest scores: of the most efficient student and of the most diligent student.

Conclusions and Considerations

This report was about the assessment of a marketing simulation exercise in terms of students' ability to develop go-to-market strategy in the classes of Master's IT students throughout three consecutive academic years – 2009/10, 2010/11 and 2011/12. A differentiation of two learning dispositions was represented – pragmatic and holistic – as a ground to determinate the dominant learning approaches of the IT students, evidenced by the learning objective pursued. Tonks's 'experiential learning' of running marketing simulations and Lin and Tu's 'ladders' of analyzing how students perceive the outcomes of BSGs use strengthened the argumentation of the simulation reasonability as a learning tool.

Master's IT students played the Marketing Simulation© pragmatically, responding to the engaging, entertaining, fun aspects of 'concrete experience', rather than holistically, thinking consistently in casual loops. They participated in *sort of* 'experiential learning', based, more or less, on 'the surface-approach' by picking out the main points of the marketing theory and practice and regarding product design, pricing, discount structure, marketing expenditures, sale-force size. Being provided with a rich array of tactical decisions, the students learned to use segment or customer needs analysis to make product design decisions and associated trade-offs. They learned to set and change prices and to appreciate the relationship between satisfaction, buying behaviour patterns and loyalty of customers and firm profitability [see Table 1].

Marketing strategies that Master's IT students developed using the simulation tool were predominantly based on "Status quo Strategy" with hybrid elements of "Increase Price Strategy". Students executed the adopted strategies by learning in pragmatic sense how to model decision casualties around some variables that they choose as constructs for successful simulation completion.

Although the Master's IT students positively accepted their involvement in experiential learning by playing the MS©, there were some open issues outlined for future administrative consideration and classroom representation. Student motivation for "learning by playing" how to develop business-to-business marketing strategies appeared to be crucial. Students needed to have a clear concept articulation of the process of "learning before and alongside playing". Hence, the instructor's role and classroom debrief sessions should strengthen their vitality of participation. The issue of "self-debriefing and learning" anticipated a cognitive empowerment of student's willingness and attitude. Instructors should incline students to think in 'ladders' derived from the simulation, and determined by sets of linkages between the attributes, consequences and values important for them. Moreover, the students had to perceive playfulness and learning performance positively because these influenced students' satisfaction and further influenced their intention to learn experientially.

References

1. Das Narayandas: Marketing Simulation: Managing Segments and Customers. Online Simulation, HBP for Educators, Copyright ©2010
2. Das Narayandas: Marketing Simulation: Managing Segments and Customers. Facilitator's Guide, HBS Publishing, Rev. 4048 (2010)
3. Kotler, P., Keller, K. L.: A Framework for Marketing Management. 3rd ed., Pearson Prentice Hall (2007)
4. Lin, Yu-Ling., Tu, Yu-Zu: The Values of College Students in Business Simulation Games: A Means-End Chain Approach. Computer and Education, vol. 58, pp. 1160-1170 (2012)
5. Tonks, D.: Using Marketing Simulations for Teaching and Learning: Reflection on an Evolution. Active Learning in Higher Education, vol. 3(2), pp. 177-194 (2002)

COMPARATIVE ESTIMATE OF THE SET OF FORMULAS FOR STRATIFIED RANDOM SAMPLES (Comparative Estimate of the set of Formulas for Stratified Random Samples)

Andreana Stoykova - Kanalieva

Abstract. The focus of this report are some problems encountered with the formulas, proposed in theory and applied in practical studies to determine the size of stratified models of representative samples. The survey of the theory and practice of representative samplings shows, that there are certain differences in using the set of formulas to determine the size of the sample in models of stratified samples. The differences are due to the proposal to plan the size of the samples by formulas according to two criteria. On the one part they are the desired maximum stochastic error of sample estimates, and on the other part – the desired mean stochastic error. It is well-known that the errors correspond to different probabilities on the security of the obtained estimates and this circumstance might cause certain problems in designing and implementing stratified samples

Keywords: sample size, random sample, stratified random sampling, standard error

The stratified models of random samples are well known in the statistical theory and practice. Their advantages have been thoroughly investigated and justified. However, the review of the theory and practice of random sample studies points out, that there are certain differences in using formulas to determine the sample size of models of stratified samples. The differences are, that the planning the samples sizes is suggested to be made according to formulas, pursuant to (two) different criteria. On the one part, it is the desired (maximum) tolerable error of the sample estimates, and on the other part - the desired standard error of these estimates. It is known, that these errors correspond to different certainty probabilities about the sizes of the actual errors of the obtained estimates and this circumstance could be associated with problems in designing and implementing stratified samples, and in interpreting the sample information. The financial costs should be added thereto, related to the formation and study of the sample - how large it could be according to these criteria.

It is also known, that the efficiency of a sample study could be searched in three directions:

1) regarding the opportunity to decrease the total size of the errors and especially the standard error;

2) regarding the cost effectiveness of the study, i.e. to cut down the costs to carry out the study, and

3) regarding the opportunity to reduce the risks of false inferences and conclusions.

This report has been written due to the following challenges.

First, the stratified models of samples are widely applied (not only independently, but also in combination of the cluster models) in the carried out random sampling by various statistical services in Bulgaria, as well as by the National Statistical Institute.

Second, some of the samplings recommended by the Statistical Office of the European Communities (EUROSTAT), are based on stratified samples, if the respective data basis is available (such as comprehensive records of economic and other units). However the methodologies elaborated and offered have not been unified and justified in sufficient details. They fail to indicate the risks of errors of interpreting the obtained estimates.

Third, the differences in the set of formulas, offered in the Bulgarian statistical literature and the formulas used by international experts cause additional problems in this respect.

Fourth, even in the specialized statistical literature (mainly Anglo-Saxon) there is no uniform use of formulas to determine the required sample size. However some of these publications neither discuss, nor propose a set of formulas to determine the required sample size, but the planning of the sample size is taken for granted.

Fifth, there is a large discussion in the statistical literature about the optimal costs of planning the sample size by strata, but the problem of linking the confidence accuracy with such costs is set aside, and also the fact, that the sample size is planned at one type of error, with one confidence level, and later the interval estimates are obtained at other types of errors, related to another confidence level.

This report attempts to clarify the above problems.

Actually the advantages of the stratified samples compared to the simple random sample could be searched in two directions. The first one refers to the standard error of the estimates. In case of a fixed sample size, the standard error of the models is decreased. The second direction refers to the fact, that with this sample model the desired accuracy could be achieved with a smaller sample size. This is a prerequisite to decrease the cost of the study to a certain extent.

Several approaches to determine the required sample size are available in the specialized statistical literature.

First, some authors like (Levy,P.,S. Lemeshow, 2008), and in the Bulgarian statistical literature the required sample size is determined in compliance with the maximum tolerable (absolute or relative) error size of the estimates. For the stratified random sample with proportional allocation the formula to determine the required sample size at given absolute size of the maximum error is presented as follows:

$$n = \frac{Nz^2 S_w^2}{N\Delta^2 + z^2 S_w^2} \qquad (4)$$

where:

z is a confidence multiplier, corresponding to a fixed probability, that the planned maximum error would not be exceeded;

S_w^2 - estimate of the variance within strata according to a fixed tested attribute[18];

[18] The variance within the strata according to a fixed tested attribute in a sample with a proportional allocation is estimated by summing the scores of the individual strata, weighted with the proportion

Δ - absolute size of the desired (planned) maximum tolerable error.

S_w^2 is estimated according to the nature of the tested characteristic features - average, relative shares or a relationship between two average sizes.

When the maximum size of the tolerable error of the estimates is fixed as a percentage size of the future estimate, the formula is as follows:

$$n = \frac{Nz^2(V_w\%)^2}{N(\Delta\%)^2 + z^2(V_w\%)^2} \quad \textbf{(2)}$$

where:

$V_w\%$ is the estimate of the coefficient of variation (the relative variance) according to a fixed attribute in percentage units

$\Delta\%$ - relative size of the desired (planned) maximum tolerable error of the estimate of the sample (or the set of estimates)

The other symbols are known.

This is the preferred way of fixing the maximum error, as usually when a sample is set up, it is expected to estimate a lot of parameters with a desired error, acceptable in practice. The sample is distributed by strata pursuant to the selected proportion criterion.

When it is a matter of stratified random sample with non-proportional allocation, called optimal sample[19], the formulas are respectively:

- in case of absolute size of the maximum tolerable error

$$n = \frac{Nz^2 S_w^2 \dfrac{1}{1+V_{S_h}^2}}{N\Delta^2 + z^2 S_w^2 \dfrac{1}{1+V_{S_h}^2}} \quad \textbf{(3)}$$

, where:

---

of the selection, i.e. $S_w^2 = \sum_h W_h S_h^2$. Here S_h^2 is the variance estimate in the individual strata, and $W_h = \dfrac{N_h}{\sum_{h=1}^{H} N_h}$ is the respective weight of the strata.

[19] When using a stratified random sample with non-proportional (optimal) allocation it is effective in view of improving the standard accuracy of the future estimates, when the coefficient of variation $V_{S_h} \geq 0{,}33$. Only in this case the standard error could be expected to drop down with at least 5 points compared to the proportional selection by strata.

V_{S_h} is the coefficient of variation (of a relative variance) of the degrees of variance by strata, i.e. of S_h

It is known that the expression $\sqrt{\frac{1}{1+V_{S_h}^2}}$ 100 shows the degree of reducing the standard error at the sampling with optimal (non-proportional) allocation of the total sample size compared to the size of the error at the sampling with proportional allocation by strata.

- in case of a maximum tolerable error set in percentages

$$n = \frac{Nz^2(V_w\%)^2 \frac{1}{1+V_{S_h}^2}}{N(\Delta\%)^2 + z^2(V_w\%)^2 \frac{1}{1+V_{S_h}^2}} \quad (4)$$

All the symbols are known here.

Second, the above approach appears also in a concealed way in the publications of M.Hansen, W.Hurwitz, W.Maddow (1953); L. Kish (1964), V. Tzonev (1971) et al, starting from the fact, that in the stratified random sample models with proportional allocation the error size is corrected by the multiplier $\sqrt{1-r^2}$, where r is the correlation coefficient within the strata between studied and stratifying attributes. Thus formulas (1) and (2) are presented as follows:

$$n = \frac{Nz^2S^2(1-r^2)}{N\Delta^2 + z^2S^2(1-r^2)} \quad (5)$$

and

$$n = \frac{Nz^2(V\%)^2(1-r^2)}{N(\Delta\%)^2 + z^2(V\%)^2(1-r^2)} \quad (\mathbf{6})$$

Here S^2 is the estimate of the total variance according to the studied attribute, and $(V\%)$ is the estimate of the respective total coefficient of variation. The other symbols are known.

Replacing r in formulas (5) and (6) with $r = \sqrt{\frac{S_b^2}{S^2}}$ and developing the expressions result in formulas (1) and (2). Here S_b^2 is the estimate of the variance between the individual strata, and S^2 - the estimate of the total variance according to a given tested attribute.

Third, in their publication P. Levy and S. Lemeshow (2008) show another way to determine the required size of the sample at stratified random sampling with proportional allocation. This is the case, when the value of the decomposition of the total variance according to a tested attribute ($S^2 = S_b^2 + S_w^2$) is known and the researcher is in a position to estimate the relationship $\gamma = \frac{S_b^2}{S_w^2}$. Then the formula for the stratified random sample with proportional allocation is as follows:

$$n = \frac{z^2 \frac{N}{1+\gamma}(V\%)^2}{N(\Delta\%)^2 + \frac{(V\%)^2}{1+\gamma}} \qquad (7)$$

The above-mentioned approaches have <u>one feature in common</u> – they make use of a desired, **maximum tolerable error** of the estimates, acceptable in practice, and function at a confidence level higher than 68%, usually 95% or 99%.

Fourth, in the methodologies, recommended now by the EUROSTAT experts for planning the required sample size under the stratified model in different tests, a set of formulas is proposed, that is borrowed by W.Cochran (1977). It needs a more detailed discussion. There <u>fundamental differences</u> appear regarding the size of the desired error. In this case the so-called **"desired variance of the sample estimates" is suggested to be used, being actually the standard error of the estimate.** It is well known, that it corresponds to 68,23% confidence level, not to exceed the actual standard error of the estimates.

The total formula to establish the sample size is:

$$n = \frac{N\sum \frac{W_h^2 S_h^2}{w_h}}{NV^2 + \sum W_h S_h^2} \qquad (8)$$

where:

w_h is the relative share of the sample size in the individual strata h, compared to the total number of the units in the respective strata in the stratified random sample with proportional allocation, i.e. $w_h = W_h = \frac{N_h}{N}$.

N_h - size of the population in the h –th strata.

N - size of the population.

For the stratified random sample with optimal allocation $w_h = W_h S_h$

V is the "desired variance of the sample estimates" as per W.Cochran. $V^2 = \frac{\Delta^2}{z^2}$

In practice this is the square of the absolute size of the standard error of the searched estimate , i.e. $V = \mu$

Afer replacing in formula (8) to determine the required sample size we obtain:

$$n = \frac{N\sum \frac{W_h^2 S_h^2}{w_h}}{N\mu^2 + \sum W_h S_h^2} \qquad (9)$$

where:

μ is the absolute error, planned for the given estimate.

It should be noted, that W.Cochran suggests also specific modifications of formula (8) to establish the sample sizes of estimation of the population total and of a relative share, in stratified random sample with proportional allocation and with optimal allocation respectively. They can not be presented here, due to obvious reasons.

The uncritical application of formulas (8) and (9) in some cases might be misleading, due to the fact, that usually when designing the sample size, the researcher sets a practically acceptable maximum tolerable error. It is determined at much higher confidence level - 95%, even 99%.

From purely mathematical point of view the planned sample sizes under formulas (1), (5) and (9) are equal. There is a difference in the actual interpretation of the estimates obtained from samples under the formulas of W.Cochran and the other authors quoted in this report. When the sample size is planned at a standard error, it means in practice, that this size has been planned so, that the estimates to be obtained with a higher error than the above-mentioned one. For example, if the size of the sample under formulas (8), respectively (9) is planned at 3% standard error, it would mean, that the estimates from this sample would be obtained with an approximate 6% error at 95% confidence level in order not to be exceeded.

However, it should be pointed out, that the formulas of W.Cochran are meaningful in practice in case, the researcher is interested in the estimates for the individual strata to be obtained with a tolerable error. Due to financial and other reasons the errors of the estimates by individual strata are planned as a rule and are obtained with higher errors than the errors of the estimates for the total population. In this case the formulas of W.Cochran, applied at

the level of the individual strata have their own justifications. The reason thereto is that having summarized the data from all the strata, the maximum tolerable error of the obtained estimates for the population as a whole will be less than the error planned for the individual strata.

Fifth, some methods and formulas to determine the sizes of the stratified models of samples are discussed and suggested in the literature, taking into account the required financing. Although being important, this approach shall not be covered in this report.

***

Without pretending to be exhaustive, this report has focused several problems, important from statistical point of view, that should be taken into consideration in the statistical practice. In general they are:

First, the set of formulas suggested in the specialized statistical literature actually solves the same problem. The differences therein are due to the differences in the approach for treating the problem.

Second, the observed differences essentially do not contradict each other, but specify different aspects of the problem.

Third, the set of formulas suggested by W.Cochran differs considerably from the others.

Fourth, the observed differences are in two directions. From the one part it is the rationale of separate formulas for the summarizing characteristics in different forms, in practice yielding the same result regarding the sample size. On the other part the researcher faces the illusion that the use of the set of formulas of W.Cochran results in a smaller error of the sample estimates, than its real size.

References

1. Джессен, Р. (1985) Методы статистических обследований, М.”Финансы and статистика” (Jessen R. - Statistical Survey Techniques)
2. Йеитс,Ф. (1965) Выборочный метод в переписях and обследованиях,М. “Статистика” (Yates. F. - Sampling methods for census and surveys)
3. Съйкова, И., Б.Чакалов (1977) Методология и методика на емпиричното социологическо изследване, С. “Наука и изкуство” (Saykova, Iv., B. Chakalov (1977) Methodology and methods of empirical sociological study, S, "Nauka i Izkustvo"
4. Цонев, В. (1970) Основи на репрезентативното изучаване, С. ВИИ ”К. Маркс” (Tzonev. V. 1970 - Fundamentals of representative sampling study)
5. Cochran, W. (1977) Sampling Techniques, J.W & Sons.
6. Hansen, M., Hurowitz, W., Madow, W. (1953) Sample Survey. Methods and Theory , J.W & Sons
7. Kish, L. (1965) Survey Sampling, J.W & Sons
8. Levy, P., Lemeshow, S. (2008) Sampling of Populations. Methods and Applications, J.W &Sons

SOME APPLICATIONS OF NONPARAMETRIC TESTS ON SURVEY DATA

Ekaterina Tosheva

Abstract: The paper treats the problems and advantages of some nonparametric tests for hypothesis testing in social and economical researches. The advantages of nonparametric tests are considered in comparison to parametric tests. Different applications of the Wald–Wolfowitz run test, the Mann-Whitney test, the Kolmogorov-Smirnov two sample test and the Kruskal-Wallis test are discussed shortly. Some of the methods are demonstrated on real data. Data are obtained from a survey on foreign investors in Bulgaria conducted at the end of 2009.

Key words: nonparametric test, survey data, foreign direct investment (FDI), Bulgaria

Introduction

In the statistical researches in social and economical sciences as well as in other areas the hypothesis testing procedure is widely used. It is known that hypothesis testing is a systematic procedure for determining whether the results from a study based on representative sample provide support for a particular theory or particular innovation, which applies to the whole population. In literature well known are a great number of statistical tests for hypothesis testing invented to solve different kind of problems. The type of the statistical test depends on the type of the sampling distribution of the test characteristic. The conventional statistical procedures are called parametric tests and are based on the sampling distribution of a population parameter (mean, variance). The purpose of parametric test is to make inferences from the sample statistic (estimation of the parameter) to the population parameter *through sampling distribution.* This distribution is completely specified and its functional form is known. Not all statistical tests are parametric. A large category of tests belongs to the so called nonparametric or distribution-free (parameter-free) tests. The term nonparametric was first used by J.Wolfowitz in 1942. According to Blalock the both terms nonparametric and distribution free are somewhat misleading, because neither such tests involve distributions without parameters, nor a population can be distribution-free. The nonparametric tests do not require the normally assumption or any other assumption which specifies exact form of the population. Nonparametric tests are concerned with problems, where the distribution functions are not specified in the extent to which they are specified by parametric tests, and where the functional form is unknown (Wolfowitz 1949). Some assumptions are also required in the nonparametric tests but they are generally weaker and less restrictive than those required in the parametric tests (Blalock 1960). Because of this, nonparametric tests are often used in place of their parametric counterparts when certain assumptions about the studyied population are questionable.

2. Main Advantages and Disadvantages of Nonparametric Tests

What are the main advantages of nonparametric tests as compared with their parametric counterparts? Firstly, parametric tests are appropriate by assumption of normal distribution of the population. But in many cases in the economical and social sciences the units from the studied populations are not normally distributed by the variable of interests.. According to the central limit theorem parametric tests work well with large samples even if the population is not normally distributed. Because the parametric tests are robust to deviations from normal distribution, so long as the samples are large the assumption of normality can be relaxed in this case. With small samples when the assumption of normality is not satisfied or is questionable the correct solution is to apply nonparametric tests. Secondly, in parametric tests data is required to be presented on interval or ratio scale. In many studies, especially in sociological and marketing surveys, the collected data are presented on the nominal and the ordinal scale. In those cases nonparametric tests are used as alternative. "Many nonparametric tests use as their data the ranks of the observation, while others are useful with data for which even ordering is impossible, i.e. classificatory data" (Siegel 1957). Furthermore the results from the nonparametric tests are less affected by outliers in the data in comparison with the results from parametric tests. At least nonparametric tests are generally easier to be applied.

However, nonparametric tests also have some disadvantages in comparison with the parametric tests. Generally, the statistical power of nonparametric tests is lower than the power of their parametric counterparts. Because nonparametric tests do not make strong assumption about the population, it is not possible to make the inference that the sample statistic is estimate of the population parameter. Nonparametric tests are better adapted to test hypotheses and they are used mostly for that purpose. Also parameter estimates are generally of more interests and more useful than hypothesis tests.(Johnson 1995). As Wolfowittz states hence the functional forms of the distribution functions are known or if there is good ground for assuming them (the situation by parametric tests) it is a loss not to make use of this information. The usage of nonparametric tests sometimes is connected with loss of precision. For example ranks are informative only about the order of the data, but discard the actual value.

Despite the abovementioned weakness of the nonparametric tests they are widely used in statistical research in all areas, and especially in sociological, marketing and behavioral studies. This paper treats the problems and opportunities of some nonparametric tests. As illustration data obtained from a survey of foreign investors in Bulgaria conducted at the end of year 2009 are analyzed using some of the discussed nonparametric tests.

3. Some Popular Nonparametric Tests

3.1. The Wilcoxon-Mann-Whitney test

The Wilcoxon-Mann-Whitney test is very popular in social sciences.[20] This test can be considered as nonparametric alternative to the t test for comparison of two unpaired groups (independent two samples t test, difference of means test). The null hypothesis in this test is that the two samples are drawn from the same continuous population (or identical

[20] This test is commonly known as either Mann-Whitney or Wilcoxon test

populations). In cases when the major differences between the two compared populations are in the central tendencies the Wilcoxon-Mann-Whitney test is one of the most powerful nonparametric tests. (Blalock 1960).

The basic principle in this test is very simple. The data from the both samples are taken and all scores are ranked from high to low, ignoring the sample membership. If the samples are less than 20 units the test statistic is U_1 (respectively U_2). To calculate it one form the samples (usually the smaller) is taken in account. For every score in this sample the number of scores in the other sample which have larger ranks is counted. The results for every case in the sample are added and the sum is the test statistic U. It can also be calculated using the formulas:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \quad (1)$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 \quad (2)$$

n_1 and n_2 are the samples' sizes in the first and second sample

R_2 is the sum of the ranks in the second sample

R_1 is the sum of the ranks in the first sample

If the samples are more than 20 units normal approximation is used. In those cases the sampling distribution of the test statistic U is approximately normal with the following parameters:

$$\bar{X}_U = \frac{n_1 n_2}{2} \text{ and } \sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \quad (3)$$

The test statistic is:

$$z = \frac{r - \bar{X}_U}{\sigma_U} \quad (4)$$

3.2. The Wilxocon Signed Ranks Test

This test is one of the most popular nonparametric tests for matched pairs and is used as an alternative of the t-test for dependent samples. As in the t test each matched pair is treated as a single case and the interest is the difference score for each pair. Applying the Wilxocon Signed Ranks Test demands to rank the differences between the scores for the pairs. This test does not require the assumption that the population is normally distributed and because of this is used when the assumption for normal distribution is questionable. It is especially useful in the case of interval scale measurement where the sample size is too small to justify the normal assumption.

The null hypothesis is that there are no differences between the scores of the both populations. As mentioned above the differences are ranked ignoring the sign of the difference. In the case of tied scores the average of the scores is given. After ranking the absolute values of differences the signs of the differences are recorded. The sums of the ranks of both the positive and negative differences are obtained. The test statistic T is the smaller of these two sums. If the samples are less than 25 units the test statistic T is the sum of the ranks of positive, or respectively the negative signs. Usually the statistic T is the smaller of these two sums. If the samples are more than 25 units normal approximation is

used and the sampling distribution of the test statistic T is approximately normal with the following parameters:

$$\bar{X}_T = \frac{n(n+1)}{4} \tag{5}$$

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} \tag{6}$$

The test statistic is:

$$z = \frac{T - \bar{X}_T}{\sigma_T} \tag{7}$$

3.3 Kruskall-Wallis Test

This test has to be considered as nonparametric alternative of analysis of variance (ANOVA) in case of independent measurements. The idea of the test is comparison between the medians of several samples (usually more than two) to determine if the samples are drawn from the same population. The scores are presented at the ordinal scale. While analysis of variance tests depend on the assumption that all populations under comparison are normally distributed, the Kruskal-Wallis test places no such restriction on the comparison. This method can be considered as extension of the Wilcoxon-Mann-Whitney test and it is used to test the hypothesis is there a relationship between two variables – one presented at the nominal scale (each category of the nominal scale variable is a separate sample) and the second presented at the ordinal scale. The idea is to compare the sums of the rankings for each of the categories of the nominal scale variable. If there are at least 5 cases for each of the categories the sampling distribution of the test statistic H is approximately X^2. The test statistic is calculated by the following formula:

$$X^2_{ем} = H_{ем} = \frac{\left(\frac{12}{n(n+1)} \sum_{i=1}^{m} \frac{\sum_{j=1}^{n_i} R_j^2}{n_i} \right) - 3(n+1)}{1 - \sum T / \left(n^3 - n\right)} \tag{8}$$

n total number of cases (sample size)
n_j number of cases in the i-th category
m number of the categories
R_j - ranks in the i-th category

$T = t^3 - t$, where t is the number of observation tied for a given rank

3.4. Kollmogorov-Smirnov Test

There are two variations of the Kollmogorov-Smirnov test – one sample and two sample test. For a single sample of data this method is used to test the hypothesis that the sample of data is consistent with a certain distribution function. In the case of two samples the method, like the Wilcoxon-Mann-Whitney test, is used to test the hypothesis that these two samples are drawn from the same continuous population (or identical populations). However, the Kolmogorov-Smirnov test can be used in cases where there are large numbers of ties, especially when the variable of interest is presented at the ordinal scale, but the data are grouped in four or more ordered categories (Blalock). The power of the test is comparable with the Wilcoxon-Mann-Whitney test and therefore this method can be considered as one of the most powerful nonparametric tests.

The idea is that if the samples are drawn from the same population the cumulative frequency distributions for the two samples have to be essentially similar. The test statistic D is the maximal difference between the cumulative relative frequencies. There are special tables with theoretical values in cases when the both samples are equal and with less than 40 units. But in _ most of the cases the following X^2 approximation is used:

$$X^2 = 4D^2 \frac{n_1 n_2}{n_1 + n_2} \qquad (9)$$

n_1 and n_2 are the samples' sizes in the first and second sample

4. Survey of Foreign Investors – Exploring Relationships per Nonparametric Tests

The above mentioned methods are used in a study of the foreign direct investors for exploring the relationships between the main characteristics of the foreign enterprises in Bulgaria and the investors' opinions about the business climate. The data are obtained from a survey of foreign investors in Bulgaria conducted at the end of 2009. The survey results can be considered as representative for medium- and large-scale foreign investors and are restricted to companies with headquarters established in Western Europe (Goev et all. 2011). Applying nonparametric tests is non-arbitrary – the variables for the main characteristics of the foreign enterprises as well as the variables for description of business climate are presented at nominal and ordinal scale and these tests as pointed above are appropriate in that case. For characterization of enterprises are used the following variables and their corresponding values:

- Sector (industry, services);
- Number of employees (up to 200, 201+)
- How long the company has been operating in Bulgaria (up to 5 years, 6-10 years, more than 10 years)
- Number of countries with company's branch offices (up to 5, 6-20, more than 20 countries)
- Taken advantage of the preferential conditions for foreign investment (yes, no)
- Type of clients (Bulgarian companies with national capital, Bulgarian companies with foreign capital and foreign companies)
- Type of the main client
- Investments in last 12 months (yes, no)

- Investments in land, buildings and equipment in last 12 months (yes, no)
- Investments in research and development (R and D), patents and licenses in last 12 months (yes, no)

For the description of the opinion of the investors the following variables are considered as appropriate:

- According to you in which way the conditions (such as legal system, taxation etc.) influence your business activities in Bulgaria (positive, negative)
- How would you estimate on the 5-scale the level of the following duties in Bulgaria – corporate tax, income tax, real estate tax, social and health insurance contributions, custom tariffs (1 – too low, 5 – too high)

The Kolmogorov- Smirnov test is used to explore whether the characteristics presented on ordinal scale (number of countries with company's branch offices and how long the company has been operating in Bulgaria) are objectively related to valuation (positive, negative) on conditions such as legal environment, taxation, financial conditions, labour remuneration, export opportunities, etc. The alternative hypothesis (that there is objective relationship) is accepted by level of significance $\alpha=0.10$ and the corresponding two dimensional distributions are shown in Table 1.

As shown in Table 1 the variable "number of countries with company's branch offices" has influence on the valuation given by the foreign investors to the legal environment, infrastructure in the country and the political climate. The variable "how long the company has been operating in Bulgaria" has significant influence only on the opinion of investors on the export opportunities.

Table1

In which way the conditions influence your business activities in Bulgaria				
	Positive	Negative	Total	KS Test Statistic and p-value
Number of countries with company's branch offices	Legal enviroment			
up to 5	11	9	20	
6-20	21	5	26	
more than 20	16	16	32	KS=5.41*
Total	48	30	78	(p=0.067)
Number of countries with company's branch offices	Infrastructure in the country			
up to 5	14	8	22	
6-20	16	7	23	
more than 20	13	19	32	KS=5.00*
Total	43	34	77	(p=0.082)
Number of countries with company's branch offices	Political climate			
up to 5	9	8	17	
6-20	3	16	19	
more than 20	6	19	25	KS=5.00*
Total	18	43	61	(p=0.082)
How long the company has been operating in Bulgaria	Export opportunities			
up to 5 years	6	4	10	
6-10 years	7	14	21	
more than 10 years	23	8	31	KS=7.15**
Total	36	26	62	(p=0.028)

** $p<0.05$, * $p<0.10$

To detect the factors with objective influence on the estimation of the burden of the duties hypothesis (presented on the ordinal scale) testing by means of the Wilcoxon-Mann-Whitney test and the Kruskal-Wallis test are provided. The first one is used when the relevant characteristics treated as factors are dichotomous, i.e. for the both values of the factor variables the corresponding estimation of the burden are considered as independent samples. Due to the fact that the samples are more than 20 units the normal approximation is used. The Kruskal-Wallis test is used in the rest of the cases, when the characteristics are on the nominal scale, (with three or more values). In the both cases the alternative hypothesis is that the relationship is objective and it is accepted by level of significance $\alpha=0.10$. The two dimensional distributions in cases of accepted alternative hypothesis as well as the estimated test characteristics are shown in Table 2 (Wilcoxon-Mann-Whitney test results) and Table 3 (Kruskal-Wallis test results).

Table 2

	Estimation of the burden of the duties						
	too low	**2**	**3**	**4**	**too high**	Total	WMW Test Statistic and p-value
Corporate tax							
Sector							
industry	1	1	1	9	4	46	
services	1	3	4		1	44	2.55**
Total	2	4	5	4	5	90	(p=0.011)
Taken advantage of the preferential conditions for foreign investment							
no	1	2	4	0	4	61	
yes	1	1	1		1	28	2.37**
Total	2	3	5	4	5	89	(p=0.018)
Investments in last 12 months							
yes	2	2	7	9	0	70	
no	0				5	20	2.85**
Total	2	4	5	4	5	90	(p=0.004)
Investments inland, buildings and equipment in last 12 months							
yes	2	9	2	7	0	60	
no	0		3		5	30	2.11**
Total	2	4	5	4	5	90	(p=0.035)
Investments in R and D, patents and licenses in last 12 months							1.67*
yes	1				0	4	

		3	0	0		4	(p=0.094)
no	1	1	5	4	5	46	
Total	2	4	5	4	5	90	
Income tax							
Investments in last 12 months							
yes		9	9	3	0	71	
no			1		1	21	2.83**
Total		0	0	1	1	92	(p=0.005)
Investments in land, buildings and equipment in last 12 months							
yes		7	2	1	0	60	
no			8	0	1	32	2.40**
Total		0	0	1	1	92	(p=0.016)
Real estate tax							
Number of employees							
up to 200	1		0	7	5	50	
201+	0		1		0	38	2.16**
Total	1	6	1	5	5	88	(p=0.031)

Table 2 (continued)

	Estimation of the burden of the duties						
	too low	**2**	**3**	**4**	**too high**	Total	WMW Test Statistic and p-value
Social and health insurance contributions							
Number of employees							
up to 200	1	6	24	14	6	51	
201+	0	10	19	11	0	40	1.72*
Total	1	16	43	25	6	91	(p=0.085)
Taken advantage of the preferential conditions for foreign investment							2.35** (p=0.020)

no	0	8	29	19	6	62	
yes	1	7	15	6	0	29	
Total	1	15	44	25	6	91	
Custom tariffs							
Number of employees							
up to 200	3	1	29	12	4	49	
201+	1	8	22	7	0	38	2.14**
Total	4	9	51	19	4	87	(p=0.033)
Taken advantage of the preferential conditions for foreign investment							
no	2	5	35	13	4	59	
yes	2	5	16	5	0	28	1.71*
Total	4	10	51	18	4	87	(p=0.087)
Investments in last 12 months							
yes	4	10	38	12	4	68	
no	0	0	13	7	0	20	1.73*
Total	4	10	51	19	4	88	(p=0.083)

** p<0.05, * p<0.10

Table 3

Estimation of the burden of the duties							
	too low	**2**	**3**	**4**	**too high**	Total	KW Test Statistic and p-value
Corporate tax							
Number of countries with company's branch offices							
up to 5	2	7	10	2	0	21	
6-20	0	9	11	5	4	29	
more than 20 countries	0	7	12	13	1	33	X^2= 6.68**
Total	2	23	33	20	5	83	(p=0.035)
Real estate tax							
Type of clients							
Bulgarian companies with foreign participation	0	2	8	12	5	27	X^2= 7.05**

Bulgarian companies with national capital	1	13	23	9	0	46	(p=0.029)
Foreign companies	0	1	10	4	0	15	
Total	1	16	41	25	5	88	
Type of main client							
Bulgarian company with foreign participation	1	1	13	8	5	28	
Bulgarian company with national capital	0	12	22	12	0	46	
Foreign company	0	1	6	5	0	12	X^2=17.05**
Total	1	14	41	25	5	86	(p<0.001)
Social and health insurance contributions							
Type of clients							
Bulgarian companies with foreign participation	0	2	15	5	5	27	
Bulgarian companies with national capital	1	12	23	12	0	48	
Foreign companies	0	2	5	8	1	16	X^2= 7.62**
Total	1	16	43	25	6	91	(p=0.022)
Custom tariffs							
Number of countries with company's branch offices							
up to 5	2	5	11	2	0	20	
6-20	1	3	15	9	1	29	
more than 20 countries	1	1	21	7	2	32	X^2= 7.60**
Total	4	9	47	18	3	81	(p=0.022)

** p<0.05, * p<0.10

The Wilxocon signed ranks test is used in order to study if there is an interaction between the given estimation of different burden and the test is provided by every couple of burden. The results from estimation of burdens are considered as dependent samples, because the same respondents (foreign investors) evaluate the different duties and the estimates of how one duty could influence the estimates of others. Because of the size of the sample z-approximation is used. The results show that there is objective relationship only between the estimation of two couples of burden: between the estimation of income tax and

real estate tax (z=1.65, p=0.098) and between social and health insurance contributions and income tax (z=1.75, p=0.080).

5. Conclusions

The application of nonparametric methods for hypothesis testing on the data from survey of foreign investors in Bulgaria shows that there are no specific category investors with entirely positive or respective negative opinion about the business environment in Bulgaria. About the fact how certain conditions influence their business activities in Bulgaria the investors with branch offices in more than 20 countries demonstrate rather negative opinions about the legal environment, infrastructure in the country and political climate. As expected, the companies which have been operating for more than 10 years in Bulgaria estimate as more positive the export opportunities in comparison to the companies with activities in Bulgaria for less than 10 years.

In terms of the burden of taxes there are certain categories of foreign investors, whose estimates on how low, respectively how high taxes are, differ objectively. Companies in the industry regard the income tax higher in comparison with those of the service sector. Companies with up to 200 employees assess as higher the property tax, the social and health insurance and the customs tariffs. In terms of the number of countries in which the companies have affiliates more favourable ratings of income tax and customs tariffs are given by companies with branch offices in up to 5 countries. As expected, income tax, social and health insurance and customs tariffs are evaluated as lower by the companies which have benefited from preferences. The companies which have made investments in the past 12 months assess as lower the income tax, the social and health insurance and the customs tariffs. Interesting is the fact that namely the companies which have made investments in the past 12 months assess the customs regime in the country negatively. In this a case it can be concluded that the reason is not in tariffs, but in other elements of the system.

The results from the conducted Wilxocon signed ranks test show that there is no special interaction between the estimation of the burden of different duties. There are only two objective relationships – between real estate tax and income tax and between income tax and social and health insurance contributions. This can be seen as evidence that the foreign investors are aware of the tax system in Bulgaria and have reasoned opinion on every tax burden.

The analysis of the relationships between the characteristics of the foreign enterprises in Bulgaria and their opinion about certain elements of the business climate is a good evidence of the use of_ nonparametric tests on survey data. Despite their disadvantages mentioned above, these methods are an appropriate alternative for exploring relationships especially in the case of nominal and ordinal scaled data as in this study.

References

1. Blalock, H., Social Statistics, McGraw-Hill Book Company (1960)

2. Chong-ho, Yo, parametric Test, http://www.creative-wisdom.com/teaching/WBI/parametric_test.shtml (2012)

3. Goev, V., V.Boshnakov, E.Tosheva, Foreign Direct Investments in Bulgaria: Main results of Foreign Investors Surveys**,** Scientific Annals of the Alexandru Ioan Cuza University of Iasi: Economic Sciences Series, LVII, p.388-410 (2011)

4. Johnson, D.H. Statistical Sirenes. The allure of nonparametrics, Ecology, 76(6), 1998-2000 (1995)

5. Siegel, S. The American Statistician, Vol.11, No 3, 13-19 (1957)

6. Wolfowitz, J. Non-parametric Statistical Inference, Berkeley Symp.on Math. Statist. And Prop., University of California Press, 93-113 (1949)

7. Съйкова, Ив., А. Стойкова-Къналиева, Св. Съйкова.Статистическо изследване на връзки и зависимости, София: Университетско издателство Стопанство (2002)

8.Тошева, Е. Статистическо изследване на зависимости, Методическо ръководсво за решаване на задачи, Издателски комплекс УНСС (2012)

Access of the Main Social Groups in Bulgaria to the Modern Computer and Communication Technologies and Services

Silviya Nikolova[1]

University of National and World Economy, Statistics and Econometrics Department, Studentski Grad,
1700 Sofia, Bulgaria
silnikolova@gmail.com

Abstract. The main goal of this paper is to be demonstrated how the new computer and communication technologies entered into the life of the ordinary people in Bulgaria and which social groups benefited firstly from these inventions without which most of the people cannot imagine their lives nowadays. The analysis examines when the members of the main social groups of the Bulgarian society firstly started to use the modern technologies and how this use changed over the period between 2004 and 2008. For each of the years, a logistic model is constructed examining the factors influencing the availability of expenditures on computers, internet or mobile technologies and services. The main data sources for the analysis are unidentified individual data from the Bulgarian Household Budget Surveys (BHBS) for the years between 2004 and 2008.

Keywords: access to computer and communication technologies and services; social classes; logistic regression

1 Introduction

The end the 20th century and the beginning of the 21st century form a period marked by a growing use of computer and communication technologies all over the world. Nowadays, people living in almost every country even cannot imagine their lives without computers or without their cell phones.

This paper presents how these technologies have become part of the life of the ordinary people in Bulgaria. What is typical for Bulgaria as a developing country is that the entering of new technologies is somehow delayed compared to the developed countries and compared to the average European level. The main reason is that during this period Bulgaria underwent a difficult transition accompanied by a sharp decrease in the living standard and high poverty rates. Usually, prices of new technologies are very high at the beginning of their lifecycle and therefore computers and cell phones were hardly affordable for a greater part of the Bulgarian population. Later on, with the stabilization of the economy, the personal computers and the mobile phones become important part of the life in Bulgaria. Nowadays, even children attending a kindergarten have their own mobile phones. If the tendencies in the entering of cellular phones and computers in Bulgaria are compared, one could notice that the use of mobile phones increased in higher speed compared to the use of computers. The main reason for this trend is the lower prices of the cell phones compared to the computer prices. Eurostat data show that the number of the subscriptions to cellular mobile services

increased from 2 per 100 inhabitants in 1998 to 107 in 2006 and 139 in 2011 [3]. This means that now almost everyone in Bulgaria uses mobile phone services and many people even have more than one subscription. This fact results mainly from the high prices of the traffic between the Bulgarian mobile operators forcing people to use more than one cell phone in order to reduce the expenditures on mobile services. Therefore, the number of the subscriptions to cellular mobile services in Bulgaria is higher than the average European level (which was 125 subscriptions per 100 inhabitants in 2011 [3]).

The same way, personal computers increased their importance for Bulgarian households during the last 20 years. However, despite the steadily increasing rate of the households owning a computer, our country is still far below the average European level. According to Eurostat data, 47% of the households in Bulgaria possessed a computer in 2011 while this rate for the 27 EU-member states was 77% [3]. On the other hand, some 22.2% of the households in Bulgaria could not afford a computer as compared to 6% at European level in 2010 [3]. This means computers are still a good which is not affordable for large part of the Bulgarian population. As regards the internet connection, the share of the households having internet access in Bulgaria increased more than four times between 2004 and 2011 (from 9.6% to 45.0%). And again, this share is still far below the European level – for the EU27 the share of the households with access to internet was 73% for 2011 [3].

2 Data and methodology

The analysis is based on unidentified individual household data from the Bulgarian Household Budget Surveys for the period 2004 - 2008. The units of observation are the households which have participated in the Bulgarian Household Budget Survey samples during the whole year. In 2004 the number of the households observed was 2846, in 2006 – 2882, in 2008 – 2871 households. The existence of expenditures on cell phones, on mobile services, on computers and on internet are the main characteristics of the households studied. The households are divided into eleven social classes accordingly to the socio-professional status of the household head [1].

Household budgets have some restrictions as regards the current research goal. Firstly, more than 50% of the households in the samples belong to the group of pensioners. Thus, the Household Budget Survey was able to indicate the start of the use of the new technologies with a relative delay because of the low income of pensioners' households and therefore postponed use of the new technologies compared to the younger groups. Secondly, the expenditures on mobile phones and mobile services were not included as a separate group in the survey until 2002. Generally, they have been existing in Bulgaria since the 90-ies but their use was low at the beginning. Thirdly, the groups of the farmers, farm workers and armed forces are underrepresented in the household budget samples (for some years there are even no households belonging to these groups in the samples).

A logistic regression models for the separate years are estimated with a dependent variable whether the household has made expenditures on mobile phones or on mobile services or on computers or on internet during the specific year. As independent variables are used the total household equivalent income, the availability up to 18 years in the household, the belonging to a specific social group (as a reference group is used the group of pensioners), the age of the household head, the type of the settlement where the household lives (the capital, cities with more than 50000 inhabitants, towns with less than 50000

inhabitants, a village. The reference group is cities with more than 50000 inhabitants), the gender of the household head and the education of the household head's partner (the reference group is a partner with secondary education).

3 Analysis

3.1 Cell phones

Table 1 presents the shares of the expenditures on cell phones, as well as the average yearly amount spent by the separate social groups. As mentioned above, there is no data about the expenditures on cell phones made by the Bulgarian households before 2004. In 2004, the share of the households with expenditures on mobile phones was 3.9%. This share increased during the following years reaching 7.3% a year after the EU-accession of Bulgaria and this increasing trend was visible for all the social groups. As expected, the groups in the upper part of the distribution were able to spend larger share of their expenditures on mobile phones. However, while the group of manual supervisors and technicians was the one with the greatest share in 2004 (10.6%), later on in 2008 the highest share was this of the service class (17,8%) . In 2004, the highest average expenditure was made in the group of self-employed – BGN 383,50 but during the next four years their average expenditures on mobile phones dropped almost twice. There was also a decrease in the average amount of the expenditures on cell phones almost in all of the rest of the groups except for the classes of unskilled manual workers and the households with a household head an inactive person. In 2008, the lowest average yearly expenditures on cell phones were made by the three groups which were mostly exposed at risk of living in poverty– farm workers, unemployed and pensioners [2].

There were no farmer's households in the sample of 2004. In 2006, their number in the sample was low and none of them made expenditures on cell phones. This type of expenditures appeared for first time in the farmers' class in 2008.

The share of the households belonging to the group of the armed forces and making expenditures on mobile phones was above the total share for all the households observed and increased over of the period. However, the high shares are most probably due to the low number of households in these groups (in 2004 they represented 0,6% of all households in the sample, in 2006 – 0,4%, and in 2008 – 0,5%). There was an increasing trend of the average amount of these expenditures too.

Table 1. Expenditures on cell phones by social classes

Social groups	2004		2006		2008	
	Share of households (%)	Average Expenditures (BGN)	Share of households (%)	Average Expenditures (BGN)	Share of households (%)	Average Expenditures (BGN)
Service (salariat)	8,9	194,82	12,4	146,94	17,8	175,75
Routine non-manual specialists	6,0	321,90	7,5	162,14	11,5	132,95
Self-employed and small scale employers	3,6	383,50	7,5	353,86	7,7	178,42
Farmers			0,0	0,00	10,3	288,97
Manual supervisors and technicians	10,6	204,80	10,0	156,48	13,8	152,30
Unskilled manual workers	4,0	119,60	6,4	152,00	3,9	243,75
Farm workers	5,6	90,00	5,6	21,00	0,0	0,00
Armed forses	17,7	177,67	33,3	87,25	26,7	211,25
Unemployed	2,0	82,00	5,6	227,73	8,0	96,86
Pensioners	1,7	196,46	3,8	124,66	3,9	75,49
Inactive	0,0	0,00	4,0	370,00	7,1	149,70
Total	3,9	211,95	6,0	156,00	7,3	131,93

Source: Autor's calculations based on the Bulgarian Household Budget Survey

The share of non-poor households with expenditures on mobile phones increased from 4.3% in 2004 to 7.7% in 2008. Among the group of the households living below the poverty line (The poverty line for 2004 is calculated as 60% of the total equivalent monthly household income. For 2008, is used the official poverty line –BGN 166 equivalent monthly income), this percentage amounted 0.8% in 2004 and increased slightly to 1.2% in 2008.

The logistic model (The results from the logistic regressions are not presented in the paper due to size restrictions, yet they are available in case of request from the author.) for 2004 measuring the probability of computing expenditures on cell phones shows that the higher equivalent household income increased the probability of buying a cell phone. On the other hand, living in a small town or in a village reduced this probability compared to living in a big city. In 2006, another statistically significant factor for increasing the probability for such expenditures appeared – the availability of children in the household. In 2008, the

greater age of the household head and the basic education of the household head's partner (as compared to a partner with basic education) were added to the negatively influencing factors. Besides, there is already a statistically significant difference among some of the social groups observed in 2008 – belonging to the salariat class, to the routine non-manual workers' class, to the group of self-employed, to the manual supervisors and technicians class or even to the group of the households with un unemployed household head increased statistically significantly the probability of buying a cell phone (as compared to belonging to the group of pensioners).

3.2 Expenditures on mobile services

Not only the expenditures on cell phones but also those on mobile services increased during the period between 2004 and 2008. Table 2 presents information about the shares of the households spending part of their income on mobile services and the average yearly amount spent by the different social groups. It is noticeable that there was a stable increase in both the share of the households paying for mobile services and in the average amount of the yearly spending on these services. The highest average expenditures were observed in the groups self-employed and the armed forces. The lowest expenditures on mobile services were observed in the groups pensioners, the farm workers and the farmers.

As regards the groups of poor and non-poor households, 11.4% of the households with incomes below the poverty line paid for mobile services in 2004 while their share reached 45.3% in 2008. However, the average amount of this yearly spending decreased between 2004 and 2008. In line with the fact that the main groups at a risk of living in poverty in Bulgaria are the unemployed, the pensioners and the families with many children [2], the households with four or more children did not make expenditures on mobile services and the share of the households with three children spending on mobile services was 62.5% in 2008 as compared to more than 90% for households with one or two children. On the other hand, for the group of non-poor households both the share and the average spending increased over the period.

The highest share of the households paying for mobile services and the highest average spending on mobile services was in the capital (73,5%), followed by those living in the big cities (71,8%), followed by those living in the small towns (66,9%), and the lowest spending was in the villages (55,2%) in 2008.

The results from the logistic regression show once again that the greater household income and the availability of children in the household increased statistically significantly the probability for paying for mobile services. Furthermore, a female household head, living in the capital or tertiary education of the household head's partner increase the chance for making expenditures on mobile services. In opposite, the higher age of the household head, living in a village (as compared to living in a city) or a partner with basic education (as compared to a partner with secondary education) reduce the probability that the household would spend money on mobile services. In 2006 and 2008, the higher education of the household head's partner was not a factor for increasing the probability for buying mobile services as compared to households with secondary education of the partner.

Table 2. Expenditures on mobile services by social classes

Social groups	2004		2006		2008	
	Share of households (%)	Average Expenditures (BGN)	Share of households (%)	Average Expenditures (BGN)	Share of households (%)	Average Expenditures (BGN)
Service (salariat)	74,8	269,75	83,7	325,25	95,2	395,51
Routine non-manual specialists	50,4	220,05	73,5	259,82	92,6	352,36
Self-employed and small scale employers	48,7	286,56	73,4	336,50	93,6	561,24
Farmers			100,0	125,00	86,2	189,29
Manual supervisors and technicians	50,9	184,38	76,4	215,33	93,5	342,08
Unskilled manual workers	28,6	134,17	57,9	181,71	81,7	242,37
Farm workers	25,0	115,43	63,9	124,26	74,2	164,94
Armed forses	70,6	305,98	91,7	540,98	100,0	579,68
Unemployed	21,5	110,26	46,7	175,42	70,8	206,02
Pensioners	14,8	112,50	29,2	140,02	46,8	154,28
Inactive	21,1	339,50	48,0	201,32	71,4	382,86
Total	28,2	180,74	47,9	208,77	65,6	268,38

Source: Autor's calculations based on the Bulgarian Household Budget Survey

3.3 Expenditures on computers

On the one hand, computers are the most expensive good among those whose consumption is studied in this paper. On the other hand, the prevailing shares of the pensioners' households in the data sets also result in small shares of the households buying a computer. In general, the share of the households spending on a computer was below 2% during the period between 2004 and 2008. The classes with the most households buying computers were the salariat and the armed forces in 2008. In contrast, among the groups of the farmers and of the farm workers there were not households with expenditures on computers. Not only the low incomes of these two groups but also the lower number of such households included in the samples could be seen as reasons for the lack of households with expenditures on a computer in the both groups.

Table 3. Expenditures on computes by social classes

Social groups	2004		2006		2008	
	Share of households (%)	Average Expenditures (BGN)	Share of households (%)	Average Expenditures (BGN)	Share of households (%)	Average Expenditures (BGN)
Service (salariat)	2,4	506,67	3,1	982,25	6,2	954,44
Routine non-manual specialists	2,8	722,14	2,4	717,71	3,0	776,56
Self-employed and small scale employers	0,0	0,00	1,1	1199,00	2,6	575,00
Farmers			0,0	0,00	0,0	0,00
Manual supervisors and technicians	2,7	987,00	3,1	1083,15	3,1	1018,46
Unskilled manual workers	0,8	920,00	0,7	1200,00	1,0	940,00
Farm workers	0,0	0,00	5,6	515,00	0,0	0,00
Armed forses	0,0	0,00	8,3	1170,00	6,7	520,00
Unemployed	0,4	200,00	1,0	830,00	0,7	1200,00
Pensioners	0,4	937,38	0,8	761,07	0,6	950,44
Inactive	0,0	0,00	0,0	0,00	3,6	750,00
Total	1,0	828,19	1,5	893,62	1,6	911,59

Source: Autor's calculations based on the Bulgarian Household Budget Survey

As expected, none of the households with incomes below the poverty line in the years observed (regardless the social group they belong to) made expenditures on computers.

As by the mobile services, the young households are those with the largest share of the households which bought a computer (4.2%) in 2008. For the middle aged households this share was 2.4% in the same year, and for the elderly households – 0.3%.

During the whole period, there were no households with three or more children which made expenditures on a computer. The largest share of the households spending money on a computer was observed in 2008 among the single child households – 4.7%. The share of the households with two children which paid for a computer was 2.3% in 2008.

The results from the logistic regression show that in 2004 the higher income and the availability of children increased the probability for making expenditures on a computer. Furthermore the tertiary education of the household head's partner also increased this probability (as compared to a household head's partner with secondary education). A female household is more likely to buy a computer in 2004 compared to a male household. The

belonging to the class of the manual supervisors and technicians increased the probability for making expenditures on a computer as compared to the belonging to the group of pensioners. The greater age of the household head, the basic education of the household head's partner, being unemployed or living in the capital or in a small town reduced statistically significantly the chance for spending on a computer. In 2008, the factors increasing the probability for buying a computer were the higher income, the availability of children in the household and the belonging to the armed forces (as compared to the group of pensioners). Compared to living in a big city all the other types of settlements reduced this probability.

3.4 Expenditures on internet access

The expenditures on internet access were included in the households' budget survey as a separate expenditure group in 2006. As could be seen from table 4, the share of the households spending money on internet connection increased rapidly between 2006 and 2008. The classes with the highest shares in 2008 were those of the salariat (59.6%) and the armed forces (60.0%). Pensioners and farm workers were the groups with the smallest share of households paying for internet connection. The increase in the average value of the expenditures resulted mainly from the increase in the prices of this type of service.

Besides the purchasing power of the households another important factor for the spreading of this service is the level of development of the exciting network and coverage of the regions. By 2008, the internet network was not well enough developed in the countryside resulting in poorly developed network in the small towns and the villages. Thus, the shares of the households with expenditures on internet were almost the same in the capital and in the big cities (35.9% and 32.3% respectively) and exceeded significantly these in the small towns and in the villages (24.2% and 7.8% in 2008). The smaller shares in the smaller settlements were result not only from the poor network coverage but also form the lower incomes of the households living there and of the greater share of the pensioners' households (especially in the rural areas).

Table 3. Expenditures on internet by social classes

Social groups	2006		2008	
	Share of households (%)	Average Expenditures (BGN)	Share of households (%)	Average Expenditures (BGN)
Service (salariat)	31,8	139,17	59,6	160,68
Routine non-manual specialists	21,8	103,71	52,2	146,25
Self-employed and small scale employers	20,2	114,24	44,9	188,70
Farmers	0,0	0,00	13,8	76,75
Manual supervisors and technicians	14,1	83,47	38,4	165,57
Unskilled manual workers	8,6	61,20	20,2	139,03
Farm workers	2,8	40,00	9,7	186,81
Armed forses	50,0	114,87	60,0	174,61

Unemployed	5,6	100,36	21,2	89,65
Pensioners	3,8	86,86	9,5	130,74
Inactive	12,0	36,67	35,7	164,98
Total	9,5	99,06	23,1	149,21

Source: Autor's calculations based on the Bulgarian Household Budget Survey

Once again, the age of the household head was expected to be one of the leading factors for making expenditures on internet in 2008. In the group of the young households 54.2% paid for internet connection while these shares were respectively 33.5% and 6.1% among the middle aged (with a household head between 30 and 64 years) and elderly households.

The number of the children in the households is also expected to be a determinant for the availability of expenditures on internet connection. In 2008, 46.7% of the households with one child and 44.2% of the households with two children spent money on internet access. In the same year, this percentage in the group of the households with three children was 18.8%. The households with four or more children in the household budgets sample do not appear to make expenditures on internet during the period between 2006 and 2008.

The share of the poor households with expenditures on internet increased from 1.2% in 2006 to 5.3% in 2008. This share in the group of non-poor households was 10.6% in 2006 and 24.2% in 2008.

The results of the logistic regression analysis show that the higher total household equivalent income and the availability of children in the household increase the probability for paying for internet. If the household belongs to the of the salariat class, to the class of the routine non-manual specialists, to the class of the self-employed and small scale employers, to the class of the manual supervisors and technicians and to the armed forces, the probability that the it would make expenditures on internet would increase statistically significantly in comparison to the households in the group of pensioners. Living in the capital increases also this probability as compared to living in the big cities in the country. A positive impact would have the tertiary education of the household head's partner too (as compared to a partner with secondary education). The higher age of the household head or the basic education of the household head's partner would decrease the chance of the household for internet access. Other factors directed to reduction of this probability were in 2006 the living in a small town or in a village (as compared to the households living in the big cities). In 2008, living in the capital or in the small towns was not any more a significant factor for having internet as compared to living in the big cities which means the difference in the network coverage decreased between these types of settlements. In 2008, a female household head was as well found to recuse the chance of having internet access as compared to the households with a male household head.

5 Conclusions

The "winners" who have benefited firstly form the new technologies were the the salariat class, the class of the routine non-manual specialists, the class of the self-employed and small scale employers, the class of the manual supervisors and technicians and the armed forces. In general, the reasons for this are the higher incomes in these groups and thus

relatively lower poverty rates among their households. The groups of the farmers, of the farm workers, of the unemployed, of the pensioners and of the households with an inactive household head benefited from the new technologies in lower speed as compared to the above stated groups.

Besides the social class, being part of the group of the poor households or living in a small town or in the rural areas would also have a negative impact on the access of the Bulgarian households to the mobile and computer technologies. Also living in a household with three or more children would decrease the probability that the households would have access to the modern computer and mobile technologies and services. The educational level of the household head's partner appears also to be one of the factors influencing the existence of expenditures on computers, internet, mobile phones and mobile services.

References

1. Тилкиджиев, Н.: Средната класа и социална стратификация, Лик, София (2002)
2. Национален статистически институт, http://www.nsi.bg
3. Eurostat, http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database

Using Software Tools for Estimation of Monthly Unemployment Rates in Bulgaria – Software Review

Alexander Naidenov

Abstract. The economic processes acceleration tendency leads to the need for faster and accurate information about the key indicators in order to observe and eventually to control these processes. The unemployment rate as one of these important indicators is usually produced on quarterly and annually basis. During the last few years Eurostat is discussing the possibilities for production of the monthly labour market indicators with the Member States. Bulgaria, as an EU Member State, have experimented the possibilities for monthly unemployment rates estimation during 2007. Considering the sophisticated estimation methodology there is a great need for the use of special software tools in order to produce reliable and precise monthly unemployment rate estimates. This paper examines the software tools available for the estimation purposes and their application in the practical problem solving situations, emphasizing on the software pros and cons in the process of the Bulgarian monthly unemployment rates estimation. The tools discussed are SPSS ver.20, Demetra+ 1.0.3, g-Calib ver. 2.0 and Ecotrim 1.01. Even though these are widely used across the EU, there are some pitfalls and issues that are considered in the Bulgarian case of estimation using data from the Labour Force Survey of the Bulgarian National Statistical Institute. The software tools screenshots are provided too.

Keywords: software, tools, estimation, monthly, unemployment, review.

1 Introduction

The unemployment rate as a key economic indicator is a subject to observation not only by the public sector but also by the private one. While a few years ago it was enough to know the overall tendency in the unemployment rate dynamics (i.e. quarter or annual basis), now dictated by the fast changing economic situation the temporal disaggregation (i.e. on a monthly basis) of unemployment time series becomes more and more indispensable. Forced by the Eurostat regulations, Bulgaria as an EU member is obliged to observe these regulations in order to achieve methodology harmonization across the EU members. Therefore there is a great need for unemployment time series disaggregation methodology improvement especially because currently data are produced on quarterly and annually basis only . Even though there were successful attempts in 2007 for the unemployment data disaggregation, now it's time to build new working tools for the estimation of the monthly unemployment rates in Bulgaria. Considering the application of those complex statistical tools it is obvious that software tools would be needed in order to facilitate the production of these estimates. As it is described below there is no single solution for the considering problem solution and this leads to the need for a mixed use of a number of software tools.

2 Monthly unemployment rate estimation approaches

The estimation of the unemployment rate in practice is based on the estimation of its two main components: number of employed persons and number of unemployed persons. The two latter constitute the rate as follows:

$$Unemployment\ rate(\%) = \frac{Number\ of\ unemployed}{Number\ of\ employed + Number\ of\ unemployed} * 100 \qquad (1)$$

Also we have to note that according to the EU practice in this monthly estimation field there are 8 main time series produced in order to provide information for 8 important demographic groups, as follows:

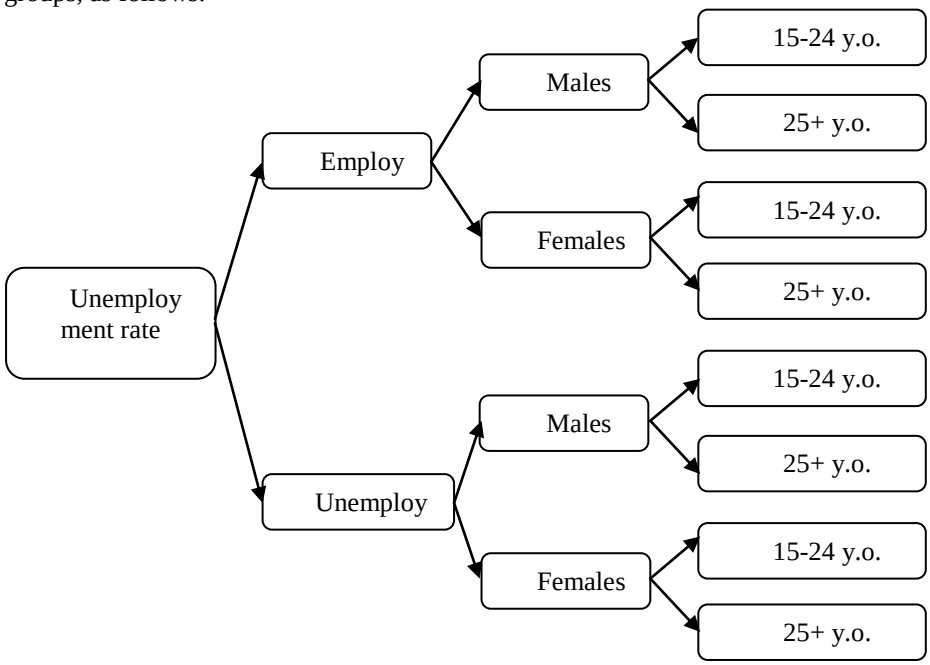


Fig. 1. Monthly unemployment rate estimation groups

For each of these groups (time series) a further elaborate statistical analysis is done by the use of two main types of estimation approaches:

Direct – using only the raw data from the quarterly based Labour Force Survey (LFS) executed by the Bulgarian National Statistical Institute (NSI).

Indirect – using LFS data and time series advanced modeling techniques (incl. interpolation and extrapolation).

We have to note here that both approaches have their pros and cons but elaborate description of the estimation methodology is not given here because the scope of this paper is dedicated to the software tools review only (a short review is provided in the next section). Also it is

important to note that the literature concerning the estimation procedures discussed here shows a great variety of methods for monthly unemployment estimation but considering the data availability and specifications in Bulgaria, the choice is limited to four possible techniques (see next section). Also the concrete results from the estimation procedures are not shown here because of the confidentiality reasons . Data used for experimentation purposes span 48 months from January 2008 until December 2011 and are provided by the NSI of Bulgaria.

3 Software estimation tools review

As it was mentioned above, because of estimation complexity, it is not possible to achieve the requested monthly unemployment rate estimates without the use of software tools. The latter are specific for each of the approach techniques and can be presented in the following table:

Table 1. Approaches, techniques and software tools for monthly unemployment rate estimation.

Approach	Technique	Software
Direct	Calibration weighting	SPSS & g-Calib
	Moving averages	SPSS & g-Calib
Indirect	Extrapolation (incl. seasonal adjustment)	Demetra
	Interpolation (data disaggregation)	Ecotrim

In the next few paragraphs a short review is provided for each technique and corresponding software used.

Technique No.1 – Calibration weighting using SPSS & g-Calib

Technique (short description): Calibration is a specific statistical technique which is mainly used to adjust the survey data to already known population totals or margins (the sum of number of employed, unemployed and out of labour force persons is adjusted to reproduce the total number of all persons aged 15+, according to the recent demographic data available). Calibration is usually based on the regression analysis and in some articles it is

stated as inverse regression. By the use of calibration we can combine the available LFS data and demographic data in order to produce the monthly estimates.

Software (short description): SPSS (Statistical Package for the Social Sciences) is a worldwide known computer program used for survey authoring and data collection, data mining, text analytics and mostly statistical analysis .

g-Calib has been developed in SPSS language by Statistic Belgium for the purposes of the data calibration. By using it we can import SPSS or MS Excel data into the software and then by the use of special g-Calib programming language we can adjust these data to the needed margins.

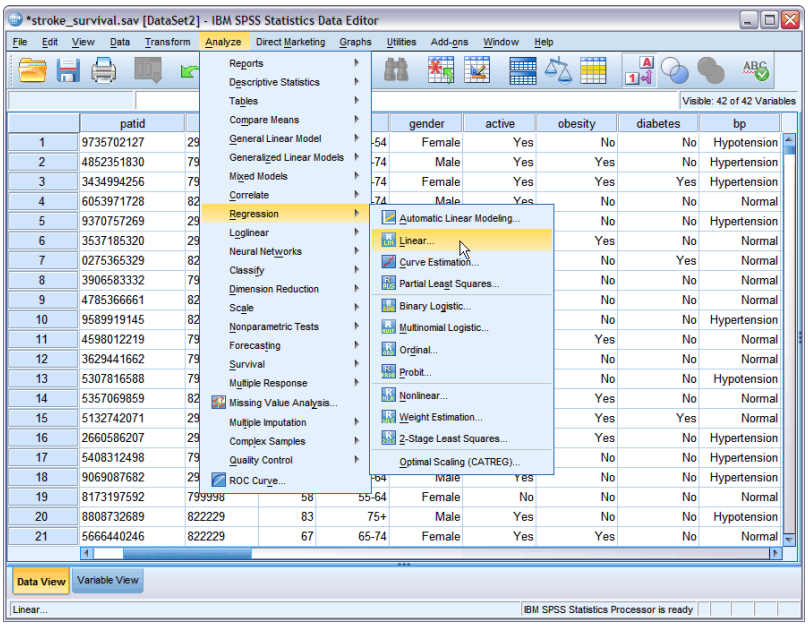


Fig. 2. Screenshot of SPSS ver.20

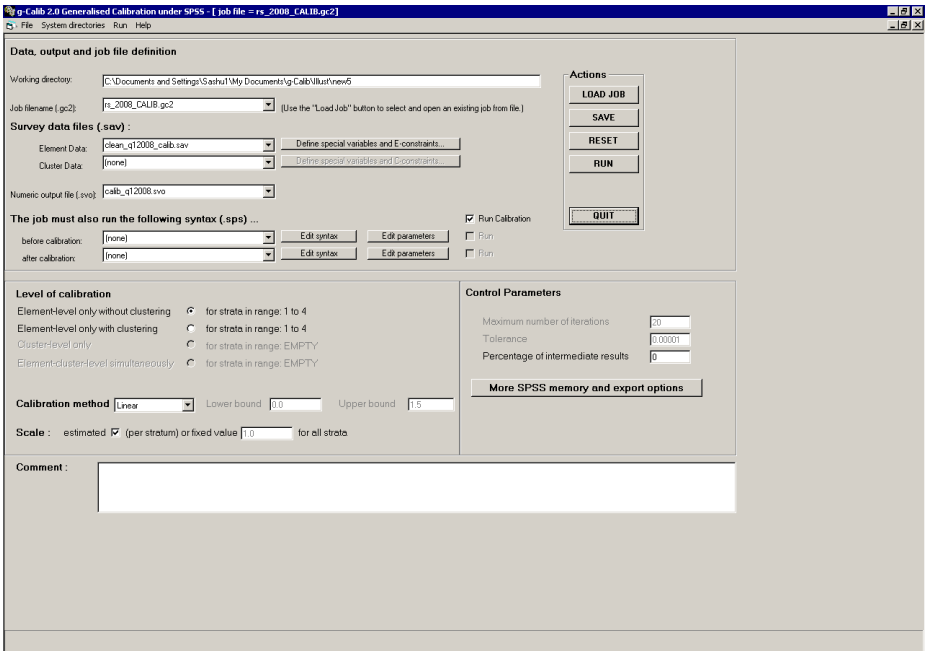


Fig. 3. Screenshot of g-Calib ver.2.0

Pros&cons: Here we consider only the pro&cons of the g-Calib only because SPSS is already well examined in many other papers and web reviews.
Pros: simple interface; semi-automated calibration process; wide functionalities (many calibration options available); uses famous data formats (.sav, .xls, etc); nicely written help file.
Cons: programming skills needed; very limited operation system compatibilities (works only on Windows 2000 SP1); too many software bugs and errors (especially concerning missing parameters or software libraries); unclear process messages; no software support available.

Technique No.2 – Moving averages using SPSS & g-Calib

Technique (short description): Moving averages concern the compilation of data from three consecutive periods e.g. months (e.g. February, March and April) and weighting these data with the last available data. Then using averaging procedures a mean estimate is used to represent the data for the middle month (e.g. March). This method is used rarely because the data production for given month requires data for the following month, so the information is produced with a great lag.
Software (short description): SPSS – see Technique No.1 for more information.

Technique No.3 – Extrapolation (seasonal adjustment) using Demetra+

Technique (short description): The presence of data for long periods of time (also called time series) gives the possibilities to analyze the data patterns if they exist. These patterns are usually analyzed by the removal of the seasonal component in those time series. The seasonal adjustment has broadened interpretation in the literature and it concerns the decomposition of the time series in four main components: seasonal, trend, cyclical and

irregular component. By the elimination of seasonality we can observe the "real" tendency in given time series. When the time series trend is known then we can predict (extrapolate) the possible future fluctuations in the data, presuming that the trend will retain in the next periods. There is a great variety of methods for seasonal adjustment but two very sophisticated but advanced algorithms are frequently used nowadays: TRAMO&SEATS (Time series Regression with ARIMA noise, Missing values and Outliers & Signal Extraction in ARIMA Time Series) and X-12-ARIMA. By the use of these methods it is convenient to decompose the unemployed and employed time series and then to extrapolate the trends in those data in order to predict the future dynamics of the derivative unemployment rate time series.

Software (short description): Demetra+ inherits the former product Demetra 2.2 and is a software tool for seasonal adjustments that was developed and published by Eurostat and implements the use of TRAMO&SEATS and X-12-ARIMA techniques. Both techniques can be divided into two main parts: a pre-adjustment step, which removes the "deterministic" component of the series by means of a regression model with ARIMA noises and the decomposition part. The two techniques use a very similar approach in the first part of the processing but differ in the decomposition part.

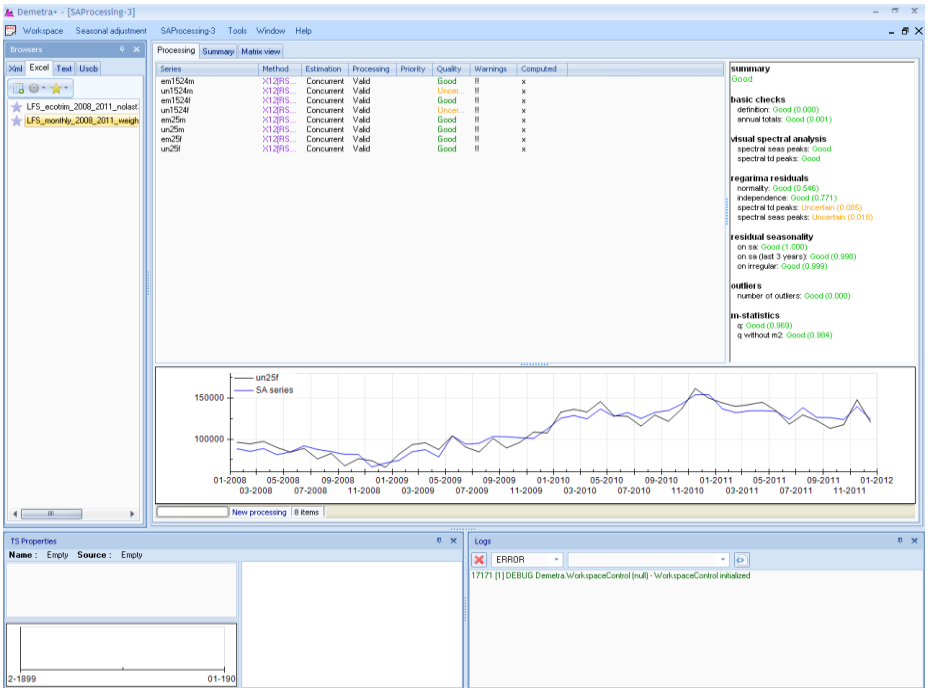


Fig. 4. Screenshot 3.1 of Demetra+ ver. 1.0.3

Pros&cons:

Pros: user-friendly interface, all Windows based OS compatibility; wide data file format usage; powerful processing and visualization engines; easy to interpret indicators and statistical characteristics; good possibilities for data and results exportation; parallel multiple

time series processing functionalities; data export possibilities; lots of well-written manuals available.

Cons: Current version 1.0.3 has a serious bug which makes it impossible to run TRAMO-SEATS; too lagged software support (it takes about 3 weeks to answer a customer's request).

Technique no.4 – interpolation (data disaggregation) using Ecotrim

Technique (short description): Often in the statistical practice it is not feasible and/or not profitable to produce data on very frequent basis e.g. on monthly instead of quarterly or quarterly instead of annually base. But sometimes pushed by need for low frequency time series data (especially for the unemployment) we need find a solution. One of these is so called interpolation technique. It is a very complex method that concerns the constructing of new data points within the range of a discrete set of already known data points also known as curve-fitting techniques using mathematical functions and statistical techniques. The interpolation is based on the regression analysis and mostly uses methods such as: random walk, Danton movement, Litterman, Fernandez univariate methods and many more methods. Because of the LFS quarterly data availability we can produce monthly estimates using interpolation techniques.

Software (short description): Ecotrim is a program for temporal disaggregation of time series developed by Eurostat.

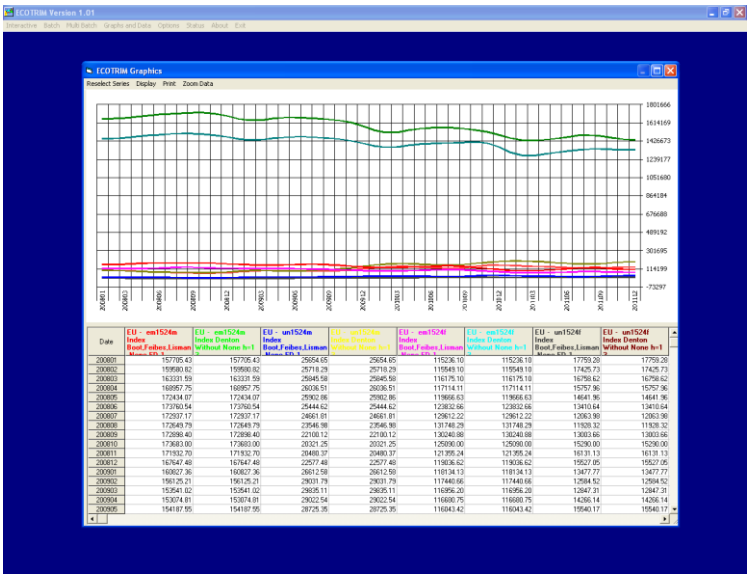


Fig. 5. Screenshot of Ecotrim ver.1.01

Pros&cons:

Pros: well-structured interactive data processing mode available; rich variety of temporal decomposition methods; standard MS Excel files usage; Windows based systems compatibility (not support for 64-bit OS version); single and multiple processing modes available; basic exportation possibilities; nicely written manual available.

Cons: No Ecotrim software development since year 2002 and therefore no updates and no support available

In the process of software usage, from statistical point of view, the estimation tools discussed here (considering their pros and cons) satisfy the needs for the monthly unemployment rates estimation. However the usage of these tools requires good knowledge not only in the field of statistics but also in computer science (esp. in operating systems and programming) which considerably narrows the software usage audience.

3 Conclusion

Summarizing the abovementioned we can conclude that the software tools used for estimation of monthly unemployment rates in Bulgaria have their positives and negatives which from one point of view give the opportunities to solve the raised estimation problems but from other point of view perpetuate the users to benefit from all the software functionalities.

Excluding SPSS and Demetra+, the software tools need serious updating not only in visual, interaction and functional aspects but also in the recent statistical methodology advances.

References

1. National Bank of Poland: Demetra+ User Manual ver.4 (2011)
2. Eurostat: ESS guidelines on Seasonal Adjustment (2009)
3. ILO: Recommendations on Seasonal Adjustment for Employment and Unemployment Data (2010)
4. Eurostat: Unemployment – LFS adjusted series (2012)
5. UN: Working Instructions for Seasonal Adjustment with Demetra (2009)
6. Infante, E.: Using DEMETRA+, Paris (2011)
7. Kocak, N.: An Analysis of German Industrial Production with DEMETRA+, Turkish Statistical Institute (2011)
8. Eurostat, http://epp.eurostat.ec.europa.eu/
9. National Statistical Institute of Bulgaria, http://www.nsi.bg/
10. Collaboration between Researchers and Official Statisticians, http://www.cros-portal.eu/page/seasonal-

Synergetic approach to economic models

Kostadin Sheyretski

University of National and World Economy

Abstract: The synergetic approach is becoming increasingly necessary for detailed study of economic processes. This paper demonstrates applying techniques from dynamical systems theory and the theory of fluctuations to economic models and makes a qualitative analysis of the equations and justifies the advantages of this approach.

Key words: synergetics, dynamical systems, mathematical economics

Introduction

The success of linear models formed the basis of the claim that linear approach for the description of the surrounding world is sufficient for analyzing all important phenomena. Complicated processes are being shattered to series of simpler ones which in turn can be described by known and well developed analytical and numerical methods. Main proof of this hypothesis is Maxwell's electrodynamics. Unfortunately this development of science, even though it's based on well-known mathematical concepts, is inapplicable for studying the world's diversity.

Henri Poincaré draws attention to the possibility of a process in which small differences in initial conditions can lead to large changes in behavior over time. This is an early reference to the discovery made in 1963 by Edward Lorenz [1]. It turns out that while a system is deterministic, ie after setting the initial conditions, its behavior is completely determined at any time, it can be unpredictable - chaotic. The evolution of a system can be observed in the state space of the system. Set of initial conditions of the dynamic (deterministic) system, which is the distance between any two points form a phase space of dynamical systems [2].

Nonlinear dynamical systems have properties that linear do not have. For example, a *non-conservative* system, ie system that is likely to lose power, can operate outside forces and fields. Such a system can generate not subsided fluctuations, the properties of which depend on the initial conditions under which the system was released. Andronov oscillators called such systems. Geometric expression of the autofluctuations in phase space is a *marginal cycle* [3].

An important concept in the theory of nonlinear systems is the *bifurcation* - acquiring a new quality of movement at least resizing. A dynamic system may have points or cycles of stable or unstable equilibrium (*steady state*), while having bifurcation can change the objects in phase space - some equilibria to disappear and to appear new ones or to be modifed the nature of the stationary state. Objects that attract trajectories in phase space are called *attractors* and the ones that repel them - repelers.

Economic Models

N.D. Kondratev created the theory of long waves which justifies the existence of cyclical economic processes [3]. The hypothesis is based on the empirical analysis of a large amount of economic indicators through different countries in the period of 100 to 150 years. Using the method of the least squares *trends* (smooth curves to which best approximates the output line) are separated, then the resulting residue averaged a nine-year moving averages. The averaging allows to smooth fluctuations with periods less than nine years. The length of the peak is measured as the distance between adjacent peaks and troughs.

Kondratev concluded that fluctuations represent deviations from process-balanced state of the economy. He believes that increasing subscribers phases of major cycles are driven by the introduction of technical innovations and development of new industries.

One of the first non-linear models in economics is proposed by Kaldor [4]:

(1) $\dot{Y} = \alpha\left[I(Y,K) - S(Y,K)\right], \dot{K} = I(Y,K) - \delta K$,

where K is the capital stock in a sector of the economy, Y - income, I - investment, S - savings and α and δ are constants. Point denotes differentiation in time. It is assumed:

(2) $\frac{\partial I}{\partial Y} > 0$, $\frac{\partial I}{\partial K} < 0$;

It is also taken into account that with supplementary conditions the tendency to investment reduction when the revenues are very volatile. Under these conditions, the system (1) can have oscillator solution.

Kaldor model is developed in discrete form

(3) $Y_{n+1} = \alpha\left[I(Y_n, K_n) - S(Y_n, K_n)\right] + Y_n, K_{n+1} = I(Y_n, K_n) + (1-\delta)K_n$.

The system of equations (3) can generate chaos.

Another non-linear model is proposed by Godwin 1967 [4], also called "class warfare", and it is the type of system is Lotka-Volterra know type. Let V be the number of production workers and the number of employers denote as U. The equations of the model are:

(4) $\dot{U} = (-a + bV)U$, $\dot{V} = (c - dU)V$,

where the constants a, b, c, and d are positive, the point is labeled with the time derivative. This model describes the continuous oscillatory nature and can eliminate its structural instability; in parentheses the first equation of system (4) is added to the function of U with negative first derivative in multitude of change U.

Many of the economic models are reduced to equations of Van der Pol [5], [6]:

(5) $\ddot{x}_1 - \mu\left(1 - x_1^{\,2}\right)\dot{x}_1 + x_1 = 0$,

where μ is a positive parameter. To construct the phase portrait of system variables are introduced:

(6) $\dot{x}_1 = x_2$,

$\dot{x}_2 = -x_1 + \mu\left(1 - x_1^{\,2}\right)x_2$.

In phase space (x_1, x_2), equation (6) has only sustainable solution in the form of marginal cycle, which can be found by numerical methods. The stationary point (0,0) is unstable, so that the closed curve winding paths inside the figure that surrounds it, but also on the curve and winding paths coming from outside of the fenced area.

In banking is used so called logistic equation [4]. Let Y_0 be an initial cash deposit in the bank rate ε .. Then the money of depositors grew by law:

(7) $Y_{n+1} = (1+\varepsilon)Y_n$.

If the bank wants to limit the growth of invested money to $Y_{\max}$, this can be done by reducing the rate proportional to the increasing amount eg

(8) $\varepsilon = \varepsilon_0\left(1-\frac{Y}{Y_{\max}}\right)$.

Comes to the famous equation

(9) $X_{n+1} = rX_n(1-X_n),\ 0 \le r \le 4,$

where $r = 1+\varepsilon_0,\ X_n = \frac{\varepsilon_0}{1+\varepsilon_0}\frac{Y_n}{Y_{\max}}$.

If $r \le 1$ only sustainable point of equation (9) is 0, for $r = 1$ zero loses stability, when $r > 1$ has two fixed points 0 and $1-\frac{1}{r}$, the first is unstable and the second stable. If $3 < r < 4$ loses stability and the second stationary point (in $r_1 = 3$), but the process starts to oscillate between the other two fixed points, $X_{n+2} = X_n$, that is occurring is a period 2 attractor. For a given value of $3 < r_2 = 1+\sqrt{6}$, a new bifurcation borns four stationary points with $X_{n+4} = X_n$, ie attractor with period 4. It was further carried a cascade of doubling bifurcation. If denotes the bifurcation parameter values of r_i, $i = 1,2,...$, it can be found the limit

(10) $\lim_{n\to\infty}\frac{r_n - r_{n-1}}{r_{n+1} - r_n} = 4.6692016...$, which is called the Feygenbaum constant.

If $r = r_\infty = 3.569944...$ behavior of the system becomes chaotic.

Lorenz equation for a small time interval can describe the dynamics of small urban systems, comprising the metropolis. [7] Considering the composition of the metropolitan urban system that does not affect the economic conditions of the metropolis, which remain structurally sound for long periods of time and can be considered stationary during the period to be considered. Local characteristics of the system are:

X - goods produced by urban system, Y - number of indigenous peoples, Z - ground rent. Production can be consumed by people or exported outside the city. Applying model

(11) $\frac{dX}{dt} = a_1(a_2Y - a_3X)$,

$$\frac{dY}{dt} = c_1\left(c_2 X - c_3 Y\right) - c_4 XZ ,$$
$$\frac{dZ}{dt} = d_1 XY - d_2 Z .$$

Parameters a_i ($i = 1,2,3$), b_j ($j = 1,2,3,4$), d_k ($k = 1,2$) are positive.

Members of $c_1\left(c_2 X - c_3 Y\right)$ and $-c_4 XZ$ determine changes in population, c_2 is interpreted as a demand for labor by firms producing unit. c_3 factor is a ratio of people choosing to work in the city's total population. The member $c_2 X - c_3 Y$ is excess demand for labor in the city. People prefer to live in places with lower rents, reflected by members $-c_4 XZ$. Member $d_1 XY$ reflects the fact that the rent positively affect X and Y.

The following modifications in the system (11) are made:

(12) $$t = \frac{t^*}{c_1 c_3},\ \sigma = \frac{a_1 a_3}{c_1 c_3},\ r = \frac{a_2 c_2}{a_3 c_3},\ b = \frac{d_1}{c_1 c_3},$$
$$x = \sqrt{\frac{c_4}{d_1}}\,\frac{d_1 X}{c_1 c_3},\ y = \sqrt{\frac{c_4}{d_1}}\,\frac{d_1 a_2 Y}{a_3 c_1 c_3},\ z = \frac{c_4 a_2 Z}{a_3 c_1 c_3}.$$

It comes to the Lorenz system

(13) $$\frac{dx}{dt^*} = \sigma\left(y - x\right),$$
$$\frac{dy}{dt^*} = rx - y - xz ,$$
$$\frac{dZ}{dt^*} = xy - bz .$$

For certain parameter values the system has chaotic behavior, and occurs in the phase space the *strange attractor*. In general dynamic system performs complex moving randomly between two *unstable focuses* - fixed points repel points describing the phase trajectory and they move in a spiral. Starting movement from any point, the system falls within unsustainable focus around which begins to move growing spiral. After a few spins the system unexpectedly goes to the second focus, where also runs a growing spiral and jumpa back to the first focus and so on. There's no movement at that frequency, the number of spins and spirals residence time near one of its focuses is entirely coincidental.

Lorenz system is dissipative. For a conservative system in which energy is constant over time, Liovil theorem applies: the phase volume is preserved, as Poincare theorem also: almost any point on any area of the phase space, moving along the trajectory will return the same area. In dissipative systems, the phase volume shrinks. From this perspective, one necessary condition dissipative system to be chaotic is:

1. phase space to be unlimited;

2. There to be a limited area in the phase space where the depicting point falls within and not random initial conditions;

3. volume sufficiently of small region of phase space with moving items constituting this field, trajectories for time t> 0 to reduce[3].

Conclusion

The article discusses the most popular models used in synergetics. Their isomorphism allows finding an application in various fields of science including economics. Nonlinear equations, although being too sophisticated analytical reviews, reflect fundamentally new phenomena in the behavior of dynamic systems. Economic processes have a complex structure, considering the influence of many factors into researching of a particular phenomenon, and there is often an inverse relationship between the variables. All this gives grounds to state that non-linear models in the future will determine the research in this area.

References

1. *Лоренц, Е.* Природата на хаоса. УИ „ Св. Климент Охридски”, 2000.
2. *Трубецков, Д. И.* Введение в синергетику: Колебания и волны. Москва. УРСС, 2011
3. *Трубецков, Д. И.* Введение в синергетику: Хаос и структуры. Москва. УРСС, 2012
4. *Панчев, Ст.* Теория на хаоса, София. АИ „Проф. Марин Дринов“, 1996.
5. *Godwin R. M.* The non-linear accelerator and persistence of business cycles. Econ. 1951. Nom. 19. PP. 1-17.
6. *Рии, Т.* Nonlinear Economic Dynamics . Lec. Notes in Econ. And Math. Systems. Vol. 336 (Springer Verlag), 1989.
7. *Занг, Б, В.* Синергетическая экономика. Время и перемены в нелинейной экономической теории. М.: Мир 1999.

The Usage of MS Excel in Mathematics Education and Linear Programming

Miroslava Ivanova[1], **Petya Valcheva**[2],

[1,2]Department of Mathematics,
Faculty of Applied Informatics and Statistics
University of National and World Economy
Sofia
[1]mi_ka@abv.bg, [2]petqv_14@abv.bg

Abstract. This article demonstrates the ease of solving the general problem of a linear programming with the use of MS Excel algorithms. The Mathematics education requires a vast knowledge of numerous formulas, mathematical methods and algorithms. Therefore, it is vital to use the appropriate software in order to illustrate their application. One of the most accessible software package is MS Office, with MS Excel being widely used. Studying Excel in depth gives us the opportunity for easy assimilation of complex algorithms and elaborate on mathematical models.Solving the assigned problems becomes faster and is easier to illustrate the basic principles and practical applications of the simplex method without bearing the burden of calculation procedures. Due to the computerized algorithm for solving the general problem of linear programming, we demonstrated to the students the practical use of the mathematical model within the economy. Alternative methods, such as Matlab, has been discussed for improving the methodology of education.

Keywords: simplex method, linear programming, MS Excel, computerized algorithm

1 Introduction

The usage of computerized methods such as MS Office (MS Excel) leads to time efficiency when compiling the algorithm and more accurate results. It also contributes to the mastery of methodology and the successful application in practical cases. This program product facilitates memorizing mathematical formulas and algorithms and gives the opportunity of better understanding of given problem.

The authers are aware of the existence of special application of MS Excel for linear optimization - Excel Solver, but the below listed algorithm presents for linear optimization solving problems efficient way of self-education and self-learning during the university mathematics course.

2 The Linear Programming Model

Each problem of linear programming involves three main elements – *decision variables*, which must be defined, *objective function* to be optimized and *constraints* that variables must satisfy as the last two are always linear functions. The *feasible region* determined by a collection of linear inequalities is the collection of points that satisfy all the constraints. Those elements from feasible region with the largest and smallest objective function value are called *optimal solutions*.

A company is able to produce *n* numbers of different goods, using for this purpose *m* numbers of defferent resources, each of which is in limited quantity b_i *(i=1,...,m).* It is known quantity a_{ij} of the resource *i* used to produce one unit of goods *j (j=1,...,n).* Let c_j be the profit, obtained from produced unit of good *j*. Task facing the manager of the company is to make a plan of production so that *total profit is maximum when limited resources*. This is the general formulation of the most decided problem.

The linear programming problem entails an optimizing process in which non-negative values for a set of decision variables $x_1,...,x_n$ are selected so as to maximize(or minimize) an objective function in the form

$$\max z = \sum_{j=1}^{n} c_j x_j \qquad (1)$$

Subject to resourse constraints in the form

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i, \ i = 1,...,m; \qquad (2)$$

$$x_j \ge 0, \ j = 1,...,n \qquad (3)$$

where a_{ij}, c_j and b_i are given constants.

The simplex method is a set of mathematical steps for solving of a linear programming problem carried out within the framework of a table called a „simplex tableau“ . In tabular form, this problem will be represented as follows (see fig.1):

C_B	B	b	c_1	...	c_j	...	c_q	...	c_n
			x_1	...	x_j	...	x_q	...	x_n
c_{s1}	x_{s1}	b_1	a_{11}	...	a_{1j}	...	a_{1q}	...	a_{1n}
⋮	⋮	⋮	⋮		⋮		⋮		⋮
c_{si}	x_{si}	b_i	a_{i1}	...	a_{ij}	...	a_{iq}	...	a_{in}
⋮	⋮	⋮	⋮		⋮		⋮		⋮
c_{sp}	x_{sp}	b_p	a_{p1}	...	a_{pj}	...	a_{pq}	...	a_{pn}
⋮	⋮	⋮	⋮		⋮		⋮		⋮
c_{sm}	x_{sm}	b_m	a_{m1}	...	a_{mj}	...	a_{mq}	...	a_{mn}
I index row	$f(X)=$		Δ_1		Δ_j		Δ_q		Δ_n

Figure 1.

The simplex tableau is the standard form of the problem (1)-(3). On the first row, above the variables x_j, are inflict the coefficients of x_j from objective function (1). Let columns B, C_B and b be formed from basic variables x_{si} (i-th equation of the constraint system (2) is determined by x_{si}, their coefficients from (1) and its values b_i *(i=1,...,m)* from (2) respectively. The columns x_j contain coefficients in front ot the homonymous variables of (2), as for the nonbasis variables are the elements of the corresponding column a_n and basis variables - elements of the corresponding single vector $(0,...,0,1,0,...,0)$, where $a_{is}=0$. In the last row of the table called *index row* are infliced the estimates of the found basic solution.

After filling the simplex tableau we follow the described below algorithm:

1. Finding of basic plan (or othe said current basic solution) and checking for *optimality*. Criterion optimality when seeking maximum is that all estimates Δ of the variables must be greater or equal to 0: $\Delta_j \geq 0$, $j=1,...,n$ *(*when seeking minimum $\Delta_j \leq 0$)

$$\Delta_j = \left(\sum_{i=1}^{m} c_{si} a_{ij} \right) - c_j, \; j = 1,...,n \qquad (4)$$

Current basic solution value can be found by the formula

$$f(x) = \sum_{i=1}^{m} x_{si} c_{si} \qquad (5)$$

2.If optimality condition is not satisfied,i.e. we derive *infeasible basic solution*, switch to a new basic plan (a neighboring vertex of the convex set), whereupon objective function value when seeking of maximum increases (if seeking minimum - decreases). The transition is accomplished by the following rules:

2.1. New basic variable x_j is chosen after comparing estimates of the variables Δ_j of the variables. In the new basic plan enter the variable, which estimate is smallest when seeking maximum (and biggest for minimum). Such variable is called *entering variable* and its column - *pivot column.*

2.2. In the process some basic variable will become *nonbasic (leaving variable).* This change of basis is done using Gauss-Jordan procedure and this is the variable, satisties the condition

$$\theta = \min\left\{\frac{b_p}{a_{pq}} : a_{iq} > 0, 1 \le i \le m\right\} \qquad (6)$$

Her row is called *pivot row* and the variable where intersect pivot row and pivot column is called *pivot element*.

If in the pivot column do not exist positive coefficient, i.e. $a_{iq} \le 0$ then the objective function is unbounded from above in the set of basic solutions (*convex polytope*) and the problem is solved. Otherwise go to the next step.

2.3. Carry out elementary transformations with the pivot element a_{pq}

2.3.1. p-th row is divided by the element a_{pq}

2.3.2. excluded x_p from the other equations of constraints system using „*rectangle rule*" . In it each new element is located by the formula:

$$a'_{ij} = a_{ij} - \frac{a_{iq}}{a_{pj}} a_{pq} \qquad (7)$$

where a'_{ij} is the value of a_{ij} in the new simplex tableau, a_{iq} is the element decumbent against the item searches in pivot column, a_{pj} is the element decumbent against the item searches in pivot row, and a_{pq} is pivot element. (see fig 3).

4 Linear Optimization in MS Excel

Using MS Excel fill the simplex tableau from the proviso of the problem. In the cell C_{m+1} specify formula (5) for the objective function value, whereas in the cell D_{m+1} define formula (4) for estimate of the variable x_1 (see Fig. 2). We fix cell references containing the coefficients in front of the basic variables from the column c_{si} and copy the formula in the whole index row till the end of the table.

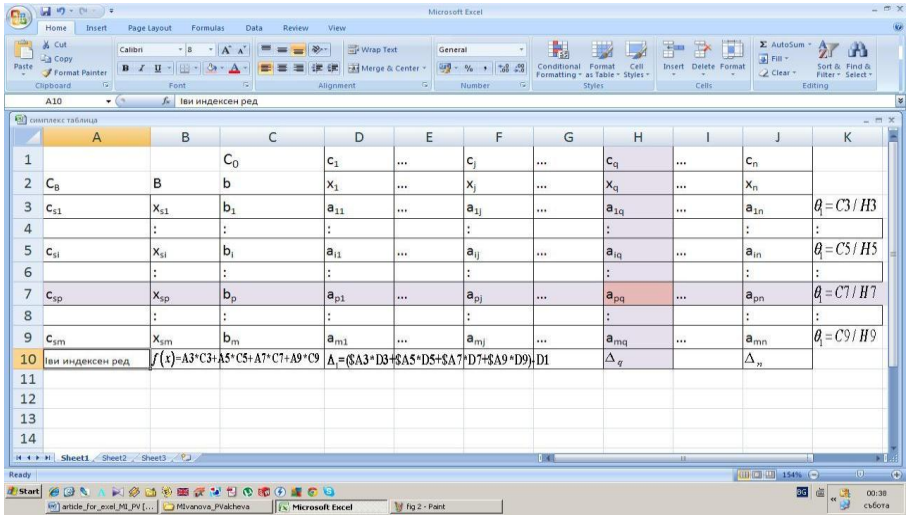


Figure 2

In order to determine the pivot column, select the range of estimates and ask to find a minimum estimate and by this way we determine „entering variable“ in the new basis. So as to determine the pivot row, i.e. „leaving variable“ from the basis, inflict the formula (6) against to every row from 1 to *m* in the column $n+1$ (in our case column *K*), define it as in *K3* and copy till the end of the table. The row against which has been obtained minimum value is pivot row (Fig. 2).

In the new simplex tableau we fill first the row which has been pivot row dividing all its elements to the pivot element (See Fig.3**).** In order to eliminate new basic variable (in our case x_q) from the rest equations of the constraints, we will use so called „rectangle rule“, shown above in formula (7), being applied to each row. We can fixate the pivot element in the denominator, because it is constant magnitude for every iteration and the element in the pivot row on whose line is currently at the moment, too. And then copy it till the end of the i-th row (See Fig. 3).

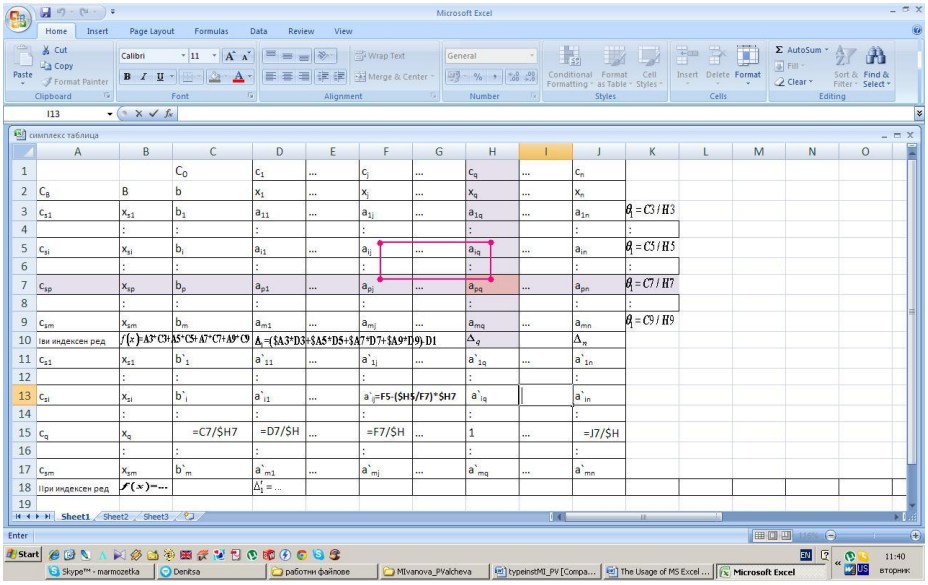


Figure 3

In the resulting new simplex tableau, in the second index row we inflict formulas for computing the estimates and in the first index row too. We repeat all this procedure until is derive *optimal solution*.

By means of MS Excel we can solve larger size problems for a short time which is only useful in mastering the material.

By next example we will demonstrate the usage of MS Exel for solving linear programming problem:

Example:
The company has three kinds of raw materials for the production of four types of production. Consumption rate by product, inventories of raw materials and income per unit of output are shown below:

How

	Product 1	*Product2*	*Product 3*	*Product 4*	*Stock*
Resource 1		*1*	*3*	*1*	*5*
Resource 2	*2*	*2*	*1*	*1*	*15*
Resource 3	*1*	*1*	*4*	*1*	*8*
Profit unit	*3*	*6*	*7*	*5*	

much production of each product to produce firm to profit from realization of production most?

To solve the given problem, we must establish mathematical model.

By x_1, x_2, x_3 and x_4 denote respectively the quantities of production of Product 1, Product 2, Product 3 and Product 4.

We seek total profit maximum, that can be expressed by objective function in following way: $\max\ z = 3x_1 + 6x_2 + 7x_3 + 5x_4$.

From the expenditure norms and the restriction in resources quantity we are able to derive constraints conditions of the problem

$$\begin{cases} x_2 + 3x_3 + x_4 \le 5 \\ 2x_1 + 2x_2 + x_3 + x_4 \le 15 \\ x_1 + x_2 + 4x_3 + x_4 \le 8 \\ x_j \ge 0,\ j = \overline{1,4} \end{cases}$$

First we convert the problem into standard form by adding *slack variables* $x_5 \ge 0, x_6 \ge 0, x_7 \ge 0$.

$$\max\ z = 3x_1 + 6x_2 + 7x_3 + 5x_4 + 0x_5 + 0x_6 + 0x_7$$

$$\begin{cases} x_2 + 3x_3 + x_4 + x_5 = 5 \\ 2x_1 + 2x_2 + x_3 + x_4 + x_6 = 15 \\ x_1 + x_2 + 4x_3 + x_4 + x_7 = 8 \\ x_j \ge 0,\ j = \overline{1,7} \end{cases}$$

There exist an source basis so we proceed to fill simplex table and solve the example using MS Excel (See Fig.4). From the resulting solution became clear that the company will achieve maximum profit if you produce 5 units of the products of the first and fourth kind.

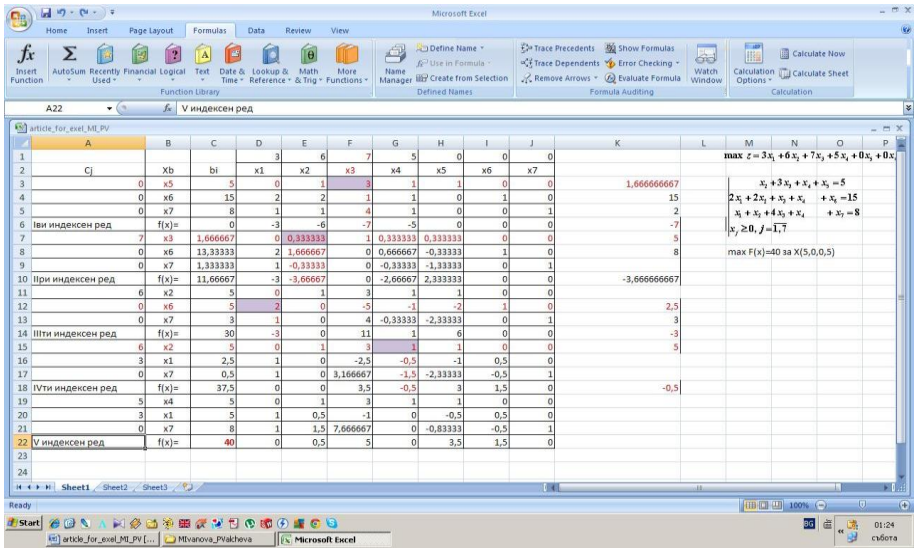


Figure 4

4 Conclusions

The main difficulties in solving "by hand" of this kind of problems comes mostly when switching to a new basic plan. Prevent mistakes that discourage students, computing is slow especially in the first meeting with algorithm the time is not enough. Therefore it is considered short problems solved with a maximum of three tables to facilitate successful mastery of the algorithm. These adapted tasks, however, do not give the opportunity to be demonstrate the practical application of the model.

The example, solved by the means of MS Excel completes in 5 tables, which are quite enough to master the necessary techniques for defining the formulas, in already known environment. The necessity to conform where exactly to position them leads to faster assimilation of the simplex method. Using MS Excel's computive capability has allowed us to solve bigger in scale problems and to show the vast the application of linear optimization.

References

1. Г. Христов, В. Кючукова, Р. Калтинска, З. Карамитева, А. Дончев, Й. Митев, П. Миланов, Н. Киров, О. Кунчев, Ръководство за решаване на задачи по математическо оптимиране, Университетско Издателство „Климент Охридски", София, 1989 г.
2. Г. Гатев, Изследване на операции, избор на решения при определеност, книга I, Издателство на Технически Университет – София, 1994 г.
3. Д. Тодоров, К. Николов, Математика, УНСС, Университетска печатница, четвърто издание, София, 2009.
4. М. Ц. Димитров, Изследване на операциите, УНСС, Университетско издателство „Стопанство", София, 2005 г.

Aspects of information security according to the main group of standards

Maria Nenova[1], Dessislava Georgieva[2],
Vesselin Gueorguiev[3] and Ivan Evgeniev[4], IEEE Members

[1] Technical University of Sofia, blvd. „Kliment Ohridski" 8, Sofia, Bulgaria,
[2] New Bulgarian University Sofia, blvd. „Montevideo" 8, Sofia, Bulgaria
{mvn, iei, veg}@tu-sofia.bg, dvelcheva@nbu.bg

1 ABSTRACT

At the beginning of the 20th century the protection of the information was necessary and has been developed in two main directions, one is to protect state secrets and the military information. So there was one reason and it is the limited access to information, lack of necessary technical equipment. The invention of computers, and later rapid development, the creation of a number of technologies for processing and expanding access to transmitted and stored data is a prerequisite for the development of programs to protect information and development of new methods and tools for this purpose. There can be differentiated two main directions. The first is the separation of users into two main groups: legitimate users are allowed to use specific data and illegitimate users for whom access to these data is prohibited.

The implementation of such protection requires processing of messages transmitted between the legitimate users to perform a cryptographic methods and tools to ensure confidentiality of information. The second strand is designed to prevent access and communication with it is implemented in such a way that carries data exchange know only legitimate users, while the rest of this fact is unknown. The role and importance of the two arms increased significantly in recent years and undoubtedly the protection of data will continue to be an important factor for development of information society.

To ensure protection in our days to the used in cryptographic systems, methods and tools the requirements become severe. They determine the improvement of existing and new cryptographic mechanisms in development. In addition to established cryptography methods, this lead to introduction of new and more complex mathematical apparatus based on number theory, theory of complexity of the calculations, information theory, etc.

All those analysis's are based on the requirements of the ISO 27 group of standards.

2 INTRODUCTION

Information security plays an important role in protecting the assets of an organisation. As no single formula can guarantee 100% security, there is a need for a set of benchmarks or standards to help ensure an adequate level of security is attained, resources are used efficiently, and the best security practices are adopted. In this paper, we give a brief introduction to the various standards and regulations that are available for information security, including ISO standards, COBIT, and etc.

Information security must become a top concern in both government [1] and business [2].

3 STANDARDS FOR INFORMATION SECURITY

3.1 ISO STANDARDS

The International Organisation for Standardisation (ISO), established in 1947, is a non-governmental international body that collaborates with the International Electrotechnical Commission (IEC)3 and the International Telecommunication Union (ITU)4 on information and communications technology (ICT) standards5. The following are commonly referenced ISO security standards:

IEC 27001:2005 (Information Security Management System - Requirements)

The international standard ISO/IEC 27001:2005 has its roots in the technical content derived from BSI standard BS7799 Part 2:2002. It specifies the requirements for establishing, implementing, operating, monitoring, reviewing, maintaining and improving a documented Information Security Management System (ISMS) within an organisation. It is designed to ensure the selection of adequate and proportionate security controls to protect information assets9. This standard is usually applicable to all types of organisations, including business enterprises, government agencies, and so on. The standard introduces a cyclic model known as the "Plan-Do-Check-Act" (PDCA) model that aims to establish, implement, monitor and improve the effectiveness of an organisation's ISMS. The PDCA cycle has these four phases. Often, ISO/IEC 27001:2005 is implemented together with ISO/IEC 27002:2005. ISO/IEC 27001 defines the requirements for ISMS, and uses ISO/IEC 27002 to outline the most suitable information security controls within the ISMS.

ISO/IEC 27002:2005 (Code of Practice for Information Security Management)

ISO/IEC 27002:2005 (replaced ISO/IEC 17799:2005 in April 2007) is an international standard that originated from the BS7799-1, one that was originally laid down by the British Standards Institute (BSI). ISO/IEC 27002:2005 refers to a code of practice for information security management, and is intended as a common basis and practical guideline for developing organisational security standards and effective management practices.

ISO/IEC 15408 (Evaluation Criteria for IT Security)

The international standard ISO/IEC 15408 is commonly known as the "Common Criteria" (CC). It consists of three parts: ISO/IEC 15408-1:2005 (introduction and general model), ISO/IEC 15408-2:2005 (security functional requirements) and ISO/IEC 15408-3:2005 (security assurance requirements). This standard helps evaluate, validate, and certify the security assurance of a technology product against a number of factors, such as the security functional requirements specified in the standard.

ISO/IEC 13335 (IT Security Management)[3]

ISO/IEC 13335 was initially a Technical Report (TR) before becoming a full ISO/IEC standard. It consists of a series of guidelines for technical security control measures.

3.2 ITIL (OR ISO/IEC 20000 SERIES)

The Information Technology Infrastructure Library (ITIL) is a collection of best practices in IT service management (ITSM), and focuses on the service processes of IT and considers the central role of the user. Since 2005, ITIL has evolved into ISO/IEC 2000021, which is an international standard within ITSM.

3.3 FIPS

The Federal Information Processing Standards (FIPS) Publication Series of the National Institute of Standards and Technology (NIST) is an official series of publications relating to standards and guidelines adopted and made available under the provisions of the FISMA30.
FIPS Publication 200, entitled *"Minimum Security Requirements for Federal Information and Information Systems"* is the second mandatory set of security standards that specify minimum security requirements for US federal information and information systems across 17 security-related areas. US federal agencies must meet the minimum security requirements defined in this standard by selecting appropriate security controls and assurance requirements laid down in NIST Special Publication 800-53 (*Recommended Security Controls for Federal Information Systems*).

4 CONCLUSION

Although there are a number of information security standards available, an organization can only benefit if those standards are implemented properly. Security is something that all parties should be involved in. Senior management, information security practitioners, IT professionals and users all have a role to play in securing the assets of an organization. The success of information security can only be achieved by full cooperation at all levels of an organization, both inside and outside.

References

[1] http://www.networkworld.com/news/2006/030706-government-cio-survey.html
[2] http://www.deloitte.com/dtt/press_release.html
[3] http://www.iso27001security.com/html/others.html

CRM System Compatible Model of Corporate Sustainability

Georgi Zabunov[1], Plamena Zlateva[2]

[1] UNWE, 1700 Sofia, Bulgaria
georgi_zabunov@unwe.bg

[2] ISER - BAS, Acad. G. Bonchev Str., Bl. 2, 1113 Sofia, Bulgaria
plamzlateva@abv.bg

Absract. The paper reviews current trends in research on corporate sustainability. It traces the emergence and development of the concept. The general concept of sustainability is combined with the vision of Triple Bottom Line. Basic sustainability trends are tracked. A model for investigating the stability of a system using Lyapunov's method is proposed. For this purpose a system of ordinary differential equations is used. Promising directions for future research are outlined.

Keywords. Corporate sustainability, model of sustainability, Lyapunov stability, system of ordinary differential equations.

1. Introduction

To define corporate sustainability is not an easy task. Even till now terms like CSR (Corporate social responsibility) and corporate citizenship are continue widely to be used. Nowadays, however, all these concepts are replaced by the broader term corporate sustainability. In the beginning it was not customary to define sustainability at corporate level.

The purpose of this report is to review the most common concepts of corporate sustainability in order to outline basic qualitative aspects of the most popular views. The ultimate goal is to find promising directions for building quantitative models of corporate sustainability at a later stage.

The most significant consideration is that these quantitative models had to be built on the basis of information available to the organization. This is information that is collected on a daily basis in corporate databases as a result of current activities.

2. Definition of Corporate Sustainability

The two main sources for precise definition of corporate sustainability are the Brundtland Commission's Report [1] and John Elkington's writings [2]. The first source defines the sustainable development as: "Development that meets the needs of the people today without compromising the ability of future generations to meet their own needs." The second source defines the sustainability at corporate level using the concept of Triple Bottom Line (TBL, 3BL or BL3). This concept is also known as the three sustainable business pillars or the three P's – People, Planet and Profit. That means that to be sustainable as a whole, the business has to be sustainable in its basic three aspects – economic (financial), social and ecologic (environmental).

We need to strongly emphasize that the economic and the social aspects of the business intersect in the field of the marketing activity of the company. As a sophisticate tool

for achieving the corporate goals through exchange, the marketing is responsible for two basic issues: developing strong company brands and building positive perceptions of the company as a whole in the society. Very often, in order to address these complex issues, marketers need help from experts outside the field of marketing. The development of sophisticate CRM systems (Customer Relationship Management Systems) is a step in this direction.

There is another field in the company's activity, where two basic aspects of the corporate sustainability intersect. The economic and the environmental aspects intersect in the field of the corporate real estate management (CREM). On the one hand, corporate real estate constitutes the main part of the company assets and has a decisive impact on the final financial results. On the other hand, buildings and building facilities are the main source of company's carbon footprint.

That is why when discussing corporate sustainability we have to pay special attention to the CRM systems and CREM (corporate real estate management) of the organization.

3. Current Corporate Sustainability Trends

The development of tools to enhance corporate sustainability remains an important task for businesses. Companies are viewing sustainability issues through a strategic lens, according to a recent survey by Ernst & Young done in cooperation with GreenBiz (Six growing trends in corporate sustainability) [3]. Respondents – executives from 24 industry sectors – indicated that the top five factors driving corporate sustainability initiatives were energy cost reduction (93%); changes in customer demand (87%); brand risks (87%); increased stakeholder expectations (86%); and competitive threats (81%). Investors and consumers are placing strong pressure on companies to consider the "triple bottom line" of environmental, social and economic performance, both in terms of the specific company and across the supply chain. The study revealed the following key observations in 2012:

- Investors will continue to press management and boards to focus on opportunities and risks related to environmental and social issues through behind-the-scenes engagement, letter-writing campaigns and shareholder proposal submissions (the most readily tracked approach). These efforts reflect the growing belief that the company's environmental and social policies correlate strongly with its risk management approach and financial performance, including in terms of supporting growth and cost-reduction opportunities.

- Regulatory changes both reflect and drive broader investor interest in proposals on environmental and social issues. In late 2009, the SEC (U.S. Securities and Exchange Commission) [4] began to allow shareholder proposals to link the term "financial risk" to discussions on the environment, among other issues, and in 2010, the agency issued guidance reminding companies of the responsibility to disclose material risks related to climate change. In 2012 the agency began to require internet service providers to include proposals asking them to be neutral (e.g., in terms of source) in how they prioritize web content; previously, the SEC allowed companies to omit these proposals from proxy statements.

- Media coverage and pending legislation are capturing investor attention on increasingly high-profile issues, such as labor conditions in the global supply chain and impacts of resource extraction practices like hydraulic fracturing. Meanwhile, more established topics – e.g., proposals to enhance corporate diversity policies, to issue sustainability reports and to adopt quantitative measures to cut greenhouse gas (GHG) emissions – continue to receive strong vote support.

- A key theme will be investor efforts to highlight the need for boards to provide improved oversight of and disclosure on the policies and procedures around political spending and lobbying activities. Furthermore, shareholder proposals on the subject will look more closely at a company's indirect political and lobbying spending through trade associations and other tax-exempt groups as these organizations may promote political or environmental agendas that conflict with the company's.

- Another key theme will be investors' growing demand for enhanced accountability and disclosure across companies' global supply chains on a full range of issues, including workplace safety, human and labor rights, and environmental practices. Shareholders are concerned that possible violations of workers' right and environmental standards – even by suppliers – may create operational, reputational and, ultimately, financial risks for the companies.

- The upcoming proxy season will also make clear investors' increasing demand for companies to apply a more integrated approach to sustainability and to make it a part of their core business strategy. One example is the emerging push for integrated reporting and for companies to address sustainability-related issues during quarterly analyst calls. In the shareholder proposal landscape, this trend is evident in proposals that seek to tie environmental and social considerations to more traditional governance issues, such as executive compensation (e.g., proposals that link pay to sustainability metrics) or the composition or the structure of the board (e.g., proposals requesting that director qualifications include expertise on environment-related issues).

4. A Model of Corporate Sustainability

The consideration of the above leads us to the possibility to develop models of the corporate sustainability. As a first step, it is more appropriate to consider the company as an autonomous system. That is not to pay attention to management impacts. In this case, the conditions of the organization (stable or unstable) are considered as a set of possible conditions. The main task is to assess the effects that lead to certain condition. Choosing two relevant indicators to characterize the organization's activity, their impact on the corporate sustainability can be addressed through a system of two ordinary differential equations. These variables can be investments and amount of dividends paid, costs of attracting new customers and costs of retaining existing customers and more. All of them allow us to build a model of corporate sustainability and the organization is seen as an autonomous system.

Here we review the tools of Lyapunov stability theory about autonomous system [5, 6]. These tools will be used to analyze the stability properties of a model of the corporate sustainability.

Consider a dynamical autonomous system

$$\dot{x} = f(x,t), \qquad x(t_0) = x_0, \qquad x \in R^n \tag{1}$$

It is assumed that $f(x,t)$ satisfies the standard conditions for the existence and uniqueness of solutions. Such conditions are, for instance, that $f(x,t)$ is Lipschitz continuous with respect to x, uniformly in t, and piecewise continuous in t.

A point $x^* \in R^n$ is an *equilibrium point* of (1) if $f(x^*,t) = 0$. It is known that an equilibrium point is *locally stable* if all solutions which start near x^* (meaning that the initial conditions are in a neighborhood of x^*) remain near x^* for all time.

The equilibrium point x^* is said to be *locally asymptotically stable* if x^* is locally stable and, furthermore, all solutions starting near x^* tend towards x^* as $t \to \infty$.

It could be assumed that the equilibrium point of interest occurs at $x^* = 0$ by shifting the origin of the system. If multiple equilibrium points exist, it is needed to study the stability of each by appropriately shifting the origin.

The equilibrium point $x^* = 0$ of (1) is *stable (in the sense of Lyapunov)* at $t = t_0$ if for any $\varepsilon > 0$ there exists a $\delta(t_0, \varepsilon) > 0$ such that

$$\|x(t_0)\| < \delta, \qquad \Rightarrow \qquad \|x(t)\| < \varepsilon, \qquad \forall t \geq t_0 \tag{2}$$

Lyapunov stability is a very mild requirement on equilibrium points. In particular, it does not require that trajectories starting close to the origin tend to the origin asymptotically. Also, stability is defined at a time instant t_0.

Uniform stability is a concept which guarantees that the equilibrium point is not losing stability. It is alleged that for a uniformly stable equilibrium point x^*, the variable δ not be a function of t_0, so that (2) can hold for all t_0. Asymptotic stability is made precise in the following definition:

An equilibrium point $x^* = 0$ of (1) is *asymptotically stable* at $t = t_0$ if $x^* = 0$ is stable, and $x^* = 0$ is locally attractive; i.e., there exists $\delta(t_0)$ such that

$$\|x(t_0)\| < \delta, \qquad \Rightarrow \qquad \lim_{t \to \infty} x(t) = 0. \tag{3}$$

As in the previous definition, asymptotic stability is defined at t_0.

Uniform asymptotic stability requires: $x^* = 0$ is uniformly stable, and $x^* = 0$ is uniformly locally attractive; i.e., there exists δ independent of t_0 for which statement (3) holds. Further, it is required that the convergence in statement (3) is uniform.

Finally, we say that an equilibrium point is *unstable* if it is not stable.

Reviewed above definitions for *stability* are *local* definitions; they describe the behavior of a system near an equilibrium point. It is said an equilibrium point x^* is *globally* stable if it is stable for all initial conditions $x_0 \in R^n$ .Global stability is very desirable, but in many applications it can be difficult to achieve.

Lyapunov's direct method (also called the second method of Lyapunov) allows us to determine the stability of a system without explicitly integrating the differential equation (1). The method is a generalization of the idea that if there is some "measure of energy" in a system, then we can study the rate of change of the energy of the system to ascertain stability.

5. Conclusion

The review clearly shows the possibilities to map fields to search for specific quantitative models of sustainability. These models have to be based on in-depth qualitative analysis of the concepts that make up the modern understanding of corporate sustainability. The discussed above is a model of this kind.

Lyapunov's method for stability analyses through a system of ordinary differential equations has many advantages:

- The necessary information is widely available in the organization;
- Using modern IT the information can be easily processed;
- It is not necessary to solve the system in order to determine the optimal values of the variables. For this purpose it is sufficient to use topological methods for determining the boundaries of system stability.

For these reasons the described model is a promising area for future research.

References

1. World Commission on Environment and Development, Our Common Future, Oxford University press, 1987.
2. John Elkington, Cannibals with Forks: The Triple Bottom Line of 21st Century Business, 1998.
3. Leading Corporate Sustainability Issues in the 2012 Proxy Season: Is Your Board Prepared? Ernst & Young, 2012, (http://www.ey.com/Publication /vwLUAssets/2012_proxy_ season/ $FILE/ 2012_proxy_season.pdf)
4. SEC sources: Staff Legal Bulletin No. 14E (CF), 27 Oct 2009; press release, 27 Jan 2010; no-action letters.
5. Parks P.E., V. Hahn, *Stability Theory*, Prentice Hall, N.Y., 1992.
6. Murray R.M., Z. *Li,* S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation*, CRC Press, 1994.